

Fighting against terrorism, engaging with Islamic science

Last week's attacks in New York and Washington were an offence against fundamental values that merits a well-targeted response, helped by science. But enhanced contacts with Islamic colleagues should also be pursued.

As *Nature* goes to press, the world is wondering how President George W. Bush, given extra powers by Congress and significant support by other nations, will respond to the barbaric killings of thousands in the United States. The impact on the scientific community has already begun to make itself felt (see page 237). Leaders of the scientific community around the world have expressed their horror and sympathy: see <http://www.nationalacademies.org>.

Science itself will play a critical role in the identification of the victims and in the unprecedented intelligence and military steps that the United States and others will now take to prevent such attacks in the future (see page 238). Many of the finest scientists and engineers will be called upon to channel their expertise into the defence of their countries against repetitions of last week's atrocity, and against its perpetrators and their defenders in every corner of the globe.

Appropriately, given last week's offence against fundamental values, most are likely to respond in full measure. A previous generation of scientists quietly helped to assure victory for the Allies in the Second World War, through the development of radar, code-breaking algorithms, and the Manhattan Project to develop the atomic bomb (the last of which, as things turned out, had the least strategic significance of the three in that conflict). This time, the challenges lie in security innovations and counter-terrorism, intelligence gathering, and enhancing an already large military advantage.

Science's role

But scientists, and others engaged with science, can do more. Last week's terrible events are utterly removed from normal relations between countries and peoples. But they are not divorced from underlying political and social forces that also affect those relations. Perhaps the least to be expected of those in a position to make a difference is some reflection on the roles of science in the cultures and societies caught up in this conflict. How might contacts between scientists and between scientific organizations, of a sort that proved valuable during the cold war, play a constructive role in long-term relations?

With thousands of dead still to be identified and put to rest, engagement of any sort will be the last thing on many people's minds. But others, deeply affected by the conflict, may feel that not to explore it could be seen as a minor victory for terrorism.

Last week's terrorist violence, after all, was not the expression of a clash of civilizations: many Islamic scholars and leaders have emphasized that the murder of the innocent is as offensive to their beliefs as to anyone else's. Their societies should not stand condemned because of extremists who disagree.

Although there could be said to be a tide of Islamic activism in the Arabic world and in Asia, there is no uniformity about it. There is a common aspect, according to knowledgeable commentators, in which resurgent Islam appears to be giving a sense of values and cultural identity to populations that may see themselves as disadvantaged or repressed within their countries. But the political contexts

and the consequences that follow are diverse — for example, only some activist groups are revolutionary in intent. Understanding that heterogeneity will be important.

Views of the Enlightenment

Differences in world view between most Western scientists and influential Islamic intellectuals (including scientists) can be profound. Societies in which Islamic beliefs are important include those actively importing Western science and technology, yet which have a distrust of the modernity and secularism of the West. Iranian political commentators, for example, saw the collapse of the Soviet system not as a triumph of the West but as a prelude to the total collapse of a system based on humanist beliefs fostered in the Western Enlightenment, which, in their eyes, committed the fatal error of divorcing a scientific understanding of nature from an appreciation of its divine aspects (see, for example, <http://web.syr.edu/~mborouje/jpr.html>). Most Western scientists, and this journal too, would consider a denial of Enlightenment values as a betrayal of everything science stands for.

In Iran and in other Islamic countries, there is no shortage of intellectual interest in the Western scientific and philosophical traditions. But questioning about the philosophical and spiritual underpinning of science can be intense. Whether only parts of Western science and culture can be imported, and whether secularization is an essential corollary of the Western Enlightenment, are important questions for Islamic scholars.

The scientist-turned-Islamic-scholar Seyyid Hossein Nasr has commented on divergent views about modern science within the Islamic world. One view, which he characterizes as 'modernist', has for over a century set about importing science without much attention to the consequences for the societies that seek to absorb it. Another view sees Western science as giving rise to ethical problems for Islam, but welcomes it nevertheless on the basis that Islam can resolve those challenges on its own ethical terms. And then there is Nasr's own view, which has been influential, and which sees science as inextricably bound up in the system of values in which it operates. It makes sense, in his terms, to identify Islamic science as related to Western science but "totally transformed into the part and parcel of the Islamic intellectual citadel" (see <http://web.mit.edu/mitmsa/www/NewSite/libstuff/nasr/nasrspeech1.html>).

Evidently, in comparison with the character of cold-war contacts, there may well be fewer common assumptions between scientific communities in the West and those in Islamic countries. There is much less knowledge of each others' scientific histories, and a consequent lack of mutual appreciation. But both inside and outside the Islamic world, there is also room for consideration of shared beliefs about the values of science, its history and its significance. Funding agencies should foster collaborations between Islamic and Western scientists and between those in the humanities studying science. Now may be a particularly good time to do so. ■





Safety first
Bringing technology
to bear in the search
for tighter security
p238



Hospital debate
Withdrawn job offer
provokes conflict-of-
interest accusations
p240



Moving on
America's National
Cancer Institute sees
its director depart
p241



Image-maker
Plankton picture
claims top spot in
photo competition
p242

Scientists react to attacks with shock and fears for the future



Washington & London

The horrific events of 11 September in New York and Washington have left scientists — in common with people from all walks of life — struggling to comprehend the brutality involved, and deeply worried about the potential consequences. Although shock and grief remain the dominant emotions, scientific leaders are already asking if the political, economic and security landscape will now shift in ways that make it impossible to adopt a 'business-as-usual' approach to research.

Scientists worldwide were quick to express sympathy and solidarity with American colleagues and with the victims of the attacks. Some scientific events were cancelled or curtailed as a mark of respect, and on 13 September the US National Academies posted on their website a selection of the messages received from other research organizations. These led with an expression of "deep sympathies and condolences" to the American people from Fathi Arafat, president of the Palestine Academy for Science and Technology. "I must say we were all utterly shocked and dismayed at the terrible human loss incurred and the excruciating pain that ensues," Arafat wrote. "May God ease your pain and grant you patience."

But scientific leaders must now think beyond their immediate emotional responses and consider the practical consequences of the current crisis. Today's scientific enterprise relies heavily on international collaboration, the free exchange of data, and unrestricted travel. In the current unstable geopolitical climate, it is unclear how each of these will be affected.

Given the nature of last week's attacks, concerns about the safety of air travel loom large (see overleaf). Leading scientists, and those in senior positions in government research agencies and high-tech industry, are among the world's most frequent fliers. There is currently a strong determination — particularly on the part of US researchers — to carry on as normal. "We should take whatever

means in security measures to mediate this so that we shake hands, look eye-to-eye, stand shoulder-to-shoulder," says Robert Goldston, director of the Princeton Plasma Physics Laboratory in New Jersey — where a meeting to celebrate 50 years of plasma physics was postponed last week. "We should not allow this event to affect the way we do science."

Even scientists who have been touched directly by last week's tragedy echo those feelings. Jan Mous, chief scientific officer with Lion Bioscience, a biotech company in Heidelberg, Germany, says that international travel will remain central to the functioning of his industry. But for Mous, such contemplation is secondary to his concern for the family of Klaus Sprockamp, Lion Bioscience's chief finance officer, who is listed among those missing at the World Trade Center in New York. "We can only pray that a miracle has happened," says Mous.

Escalating tension

The extent to which scientists will resume their hectic travel schedules is likely to hinge on future events, however. Any suggestion that airliners remain vulnerable to hijacking by suicide bombers is bound to dent people's resolve to keep flying. Already, carriers are reducing the number of scheduled flights. And if last week's outrage leads to war between the United States and scattered terrorist groups and the states that harbour them, as seems likely, international scientific travel could be severely curtailed. "If there is a riposte, then I stop travelling immediately and completely," the research director of a leading European pharmaceutical company told *Nature*.

Such concerns are now foremost in the minds of the organizers of major scientific meetings. In the days immediately after the attacks, organizers of the largest pending conferences were optimistic that they could proceed as normal. But as events continued to unfold, some scientific societies began to reconsider their plans.

The American Society for Microbiology, for instance, which was expecting some 15,000 delegates to attend its 41st Inter-science Conference on Antimicrobial Agents



Devastation: workers remove rubble after the collapse of the World Trade Center.

and Chemotherapy in Chicago on 22–25 September, initially issued the following statement: "We believe that in response to the national tragedy that has occurred, holding this important scientific meeting is in the best interests of human health and welfare." But by 14 September, the conference's organizers had decided to postpone the meeting until December, noting that "uncertainties concerning the availability of safe, scheduled air travel force us to take this action in

AP the best interests of our infectious-disease community”.

While the organizers of future conferences reviewed their plans, other meetings were disrupted by last week's interruptions to air travel. The Cardiovascular Research Foundation cut short a meeting in Washington that opened on the day of the attacks, and the American Association of Cardiovascular and Pulmonary Rehabilitation's annual meeting in Minneapolis, which was due to start on 12 September, was abandoned. Across Europe and Asia, many meetings went ahead without American participants.

As concerns grow about escalating global instability and the likelihood of the United States launching a military response, the ramifications for science could run much deeper than the disruption of conferences.

An increased focus on security seems inevitable. Last week, for instance, the US Senate Committee on Governmental Affairs was told that federal agencies, including those responsible for research, have given inadequate thought to protecting their computer infrastructure against terrorist and other threats.



Research agencies may now be asked to tighten their security procedures across the board. It is also possible that policies on the exchange of scientific data will be reviewed. “At present, we distribute and share scientific information without regard to where it's going,” says Graham Cameron, joint head of



the European Bioinformatics Institute in Hinxton, near Cambridge, UK. But he speculates that the US government and its allies could demand that certain countries are excluded from access to a range of scientific data — such as the genome sequences of pathogenic microorganisms.

Technology will assist the fight against terrorism



William Triplett, Washington

In the immediate aftermath of last week's attacks, attention has focused on the importance of beefing up the human element in counter-terrorism. The need for better-trained security

personnel and a stronger emphasis on the human aspects of intelligence-gathering have been cited repeatedly. But experts stress that science and technology will also be fundamental.

Take airport security, for example. US airlines have been heavily criticized for using poorly trained employees to screen passengers and baggage. But aviation specialists say that, even with outstanding staff, improved technologies are needed to maximize security. “With excellent people you can build an aviation-security system that can be very good,” says Billie Vincent, a security consultant based near Washington's Dulles airport, who is a former director of civil-aviation security for the US Federal Aviation Administration (FAA). “But you can build a better security system if you have excellent people and then integrate the right technology.”

Many of these technologies are already available, at least in prototype form, and their application could now be seriously

considered. They include ‘biometrics’ for confirming the identity of passengers, computerized pattern recognition to identify weapons hidden in baggage or under clothing, and even an extension of auto-pilot technologies to prevent hijackers from assuming manual control of airliners.

“The strategy used to be keeping bombs and weapons off airplanes,” says Rick Charles, a specialist in aviation security at Georgia State University in Atlanta. But facing a new breed of terrorist prepared to take their own lives by turning airliners themselves into weapons, “now we need to look at how to keep the wrong people off airplanes”.

The field of biometrics — which recognizes individuals by fingerprints, iris patterns and so on — has a role to play, argues Charles. “If you had sophisticated identification devices at ticket counters, and they were linked to FBI, Interpol and CIA databases, it would certainly help.” A more comprehensive system would require all passengers to carry biometric identity cards, which could be checked and verified.

Some technologies, such as enhanced X-ray machines that can detect objects hidden under clothing, may now win public acceptance that was previously elusive on grounds of privacy. American Science & Engineering, based in Cambridge, Massachusetts, has a device called a Z Backscatter X-ray, which, when used in

conjunction with conventional X-rays, offers sharper image resolution and can discriminate between different materials — pinpointing explosives, for example. The equipment is used by US prisons, but the FAA has yet to certify it.

Experts argue that such machines, allied with computer-image-recognition systems, are needed to tighten airport security, however well-trained security personnel may be. People simply cannot maintain the levels of concentration to reliably spot suspicious objects on the screens of X-ray scanners, says Douglas Harris, chairman of Anacapa Sciences, a security-systems-analysis company in Santa Barbara, California. “Vigilance drops within minutes,” he says.

A software package called Threat Image Projection, developed by PerkinElmer Instruments of Boston, Massachusetts, is capable of identifying weapons of various shapes from X-ray images. It has been tested at Atlantic City airport in New Jersey and is being deployed elsewhere. But its pattern-recognition algorithms would need to be modified to identify the knives and box-cutters reportedly used by the terrorists responsible for last week's atrocities.

There is also scope to improve security on board aircraft. Strengthening cockpit doors with Kevlar, a lightweight, bulletproof substance, for example, would cost only \$2,000 per aircraft — and could be combined



Science officials in the regional neighbours of Afghanistan — the most likely target of US military action — are acutely aware of the potential for a security clamp-down to affect international scientific collaboration. “It is too early to say in what form restriction will come,” says Ragunath

Mashelkar, secretary to India’s Department of Scientific and Industrial Research. “But there is bound to be some restriction on the freedom that non-Americans currently enjoy in the United States.”

More generally, the crisis could have a profound effect on the resources that are

made available for research. The US federal budget will come under pressure as money is released for the immediate relief effort, for eventual rebuilding at the sites of the attacks and to finance the expected military action. Already, the US Congress has authorized the spending of \$40 billion for rebuilding and to step up security. “The money has to come from somewhere,” observes Robert Eisenstein, assistant director of the mathematical and physical sciences directorate of the National Science Foundation.

Although officials such as Eisenstein are duty-bound to consider the potential implications of the crisis for their agencies’ budgets, such concerns are not foremost among the scientists who spoke to *Nature* over the past week. “There are thousands of widows, widowers and orphans,” says Princeton University physicist William Happer, director of the Office of Energy Research in the US Department of Energy during the early 1990s. Under the circumstances, he says, most scientists will accept that fundamental research is not going to be the US government’s top priority. ■

Reported by Alison Abbott, David Adam, Josette Chen, Rex Dalton, K. S. Jayaraman and Paul Smaglik.



with procedures to keep cabin doors locked during flight. US officials are also discussing a major expansion to the Air Marshal programme, to put armed guards on many more scheduled flights. Discharging a firearm during pressurized flight is dangerous — but Israel’s airline, El Al, equips its guards with nylon composite bullets that can disable hijackers without threatening cabin walls.



Detection aids: could an experimental explosives-detection portal (left) augment existing X-ray screening equipment at airports?

Charles also believes that auto-pilot technology could be further developed to prevent hijackers seizing control of aircraft. “We’ve had fully automatic landing capability for a long time.” With current technology, he says, “it wouldn’t take great additional effort to implement a system that would activate fully automatic flight and landing and disable all manual control inputs”.

But the system could create problems if it malfunctioned, and denied control to legitimate pilots. Developing telemetry to override manual inputs and assume control from the ground in the event of hijacking would help. But that, in turn, might be vulnerable to someone hacking into the system. Such complications illustrate how, in combating terrorism, securing one area of vulnerability often creates another.

Some researchers complain that the FAA’s lengthy certification process has delayed the adoption of promising security techniques. A

team at Sandia National Laboratories in Albuquerque, New Mexico, for example, developed a screening portal in 1997 that can detect explosives carried by passengers. After two years of tests at Albuquerque’s airport, a New Jersey company, Barringer Technologies, bought rights to the technology, but the FAA has yet to certify it.

The FAA will now come under pressure to expedite its certification procedures. Its research priorities may also be reviewed. The administration spends around \$50 million a year on research and development, about \$40 million of it on explosives detection.

However, the events of 11 September suggest that, whatever technologies become available, they cannot provide watertight security. There is no “magic bullet”, warns John Hansman, a physicist and director of the International Center for Air Transportation at the Massachusetts Institute of Technology. “Any technical solution for security will be highly vulnerable to a well-planned attack.”

As the US authorities investigate the vulnerabilities that allowed last week’s attacks to occur, a top priority will be providing better intelligence on terrorist activities. “Intelligence tells you what the threat scenario is,” says Harris, “but intelligence on terrorism is a severe weakness of the system right now.” Addressing that weakness may require vast reforms of both human spying and the use of technologies such as electronic communications interception. ■

Additional reporting by Rex Dalton.

Job refusal sparks row over mind-drug critic

David Spurgeon, Montreal

Canada's leading research university is coming under fire after one of its affiliated hospitals withdrew a job offer to a prominent critic of psychological drugs.

Late last year, the Centre for Addiction and Mental Health (CAMH) at the University of Toronto hastily retracted an offer to David Healy, a reader in psychological medicine at the University of Wales College of Medicine in Cardiff, to become clinical director of its mood and anxiety disorders programme.

In a letter to Robert Birgeneau, president of the University of Toronto, the contents of which were made public earlier this month, 27 eminent scientists said that the handling of the affair was "besmirching the name of one of North America's great research universities", and was an "affront to the standards of free speech and academic freedom".

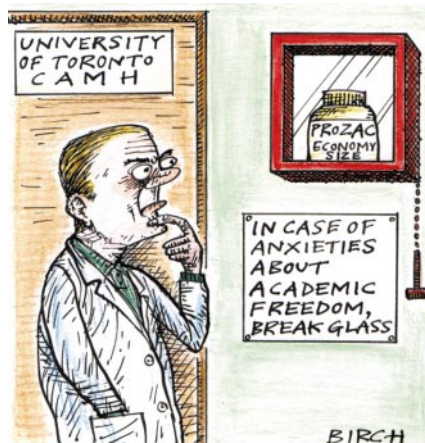
David Naylor, the university's dean of medicine, hit back on 7 September, calling the letter "misdirected" and saying that the authors — including two Nobel prizewinners and six past presidents of professional societies — should have discussed its contents with the university before releasing it publicly.

The post offered to Healy, which included a 'status-only' professorship, was rescinded after he gave a lecture at the university in which he was said to have criticized the use of so-called 'psychotropic' drugs. These include the antidepressant Prozac, made by Eli Lilly and Co., which contributes funds to the CAMH. Media reports in Canada suggested that this link with industry influenced the decision to withdraw the employment offer.

Naylor denies any such influence. "This is an issue where a partner (CAMH) has made a decision on a clinical leadership position and then, unfortunately, decided to reverse the decision," he says. He adds that the university will still provide an academic appointment



Focus of controversy: the Centre for Addiction and Mental Health in Toronto.



for Healy, provided that he gets a medical appointment in an affiliated hospital.

Naylor says the recruitment process has been "less than ideal", but argues that criticism of the university "reflects a fundamental

misunderstanding" of its relationship with clinical affiliates, such as the CAMH. The offer to Healy involved no university resources and was not a tenure-track appointment, Naylor says.

Those who heard the lecture, on 30 November, differ about what Healy actually said. Paul Garfinkel, head of the CAMH, says that CAMH staff present told him they were "shocked and alarmed" at Healy's statements on the hazards of some antipsychotic drugs. Garfinkel says that Healy has lost the support and respect of colleagues.

"This has nothing to do with academic freedom," says Garfinkel. "No one disputes Healy's freedom to say whatever he wants, and in a university setting this is encouraged. But in a hospital we have to worry about quality of patient care first."

Healy says that he did not say anything new in his lecture, and that it was well received by the audience, including many clinicians. This is "in complete contrast to what has been said by Dr Naylor and Dr Garfinkel", he adds. Healy also says that few CAMH staff were present, and that the only objector was CAMH chief physician David Goldbloom, whom Healy says was not enthusiastic about what he said on the hazards of some antipsychotic drugs.

Thomas Ban, emeritus professor of psychiatry at Vanderbilt University in Nashville, Tennessee, who composed the letter to Birgeneau, says that academic freedom is the sole issue involved. Margaret Somerville, director of the Centre for Medicine, Law and Ethics at McGill University in Montreal, says: "When you put universities and industry together, you get these conflicts."

♦ www.pharmapolitics.com

Mad cow disease comes to Japan

Jim Giles

The reach of mad cow disease extended around the globe on 10 September, when Japan announced the first confirmed case of bovine spongiform encephalopathy (BSE) in a cow born outside Europe.

Japanese officials say that a cow in the Chiba prefecture, east of Tokyo, has been diagnosed with the disease. All previous cases outside Europe involved cattle imported from Britain.

Researchers think that BSE originally entered Japan through infected meat and bonemeal cattle feed imported from Britain during the early 1990s. But worryingly for Japan's farmers, the disease

may now have spread to domestic cattle feed.

Unconfirmed reports say that the animal that died was 5 years old. Experts say that British cattle feed was probably BSE-free by the time the cow was born, suggesting that the infection came from foodstuffs produced in Japan itself, perhaps using animal parts from cattle that had carried earlier, undetected infection from British feed.

"There could have been a new round of infection," says Marcus Doherr, a BSE expert at the University of Bern in Switzerland. But Doherr says that Japan is unlikely to see an epidemic on the British scale, as cattle feed made from animal parts was not as widely used in Japan as in Britain.

Cancer institute director's exit leaves NIH in the lurch

Matthew Davis, Washington

Richard Klausner resigned on 11 September as director of the National Cancer Institute (NCI) to become president of the new Case Institute of Health, Science and Technology. The institute is bankrolled by Steve Case, the founder of America Online (AOL).

Klausner's departure after six years at the NCI accentuates the leadership void at the US National Institutes of Health (NIH). The biomedical research agency now lacks permanent directors both at its headquarters and at its largest institute — the NCI accounts for \$3.7 billion of the NIH's \$20.4 billion budget this year.

The positions are vacant at a time when the agency is facing several vexing issues, including the lack of a long-term strategy for managing its recent unprecedented expansion. There is also controversy over the government's role in funding human embryonic stem-cell research and growing concern about patient safety in NIH-supported clinical trials.

Sources inside and outside the NIH had been predicting Klausner's imminent departure for more than a month. They said Klausner and officials at the Department of Health and Human Services, the NCI's parent agency, were at odds over salary increases Klausner had given senior administrative staff and over travel by NCI staff to scientific meetings.



Outward-bound: Klausner wants to be more directly involved with research at the bench.

Klausner says that such issues "were not even a component" of his decision to resign. He attributes speculation about tensions between himself and health-department officials to "the internal rumour mill".

"Whenever an administration changes, these things come and go," he says. "But they were never in my mind as I was determining whether I should stay or go. There is a tendency for people to put forth their own issues when they're trying to explain events."

Klausner adds that he had developed a "good relationship" with the Bush administration. He says that health-department secretary Tommy Thompson had asked him "numerous times" whether he was interested in moving from the NCI to the NIH directorship. He had declined every time, he says, but adds that the discussions never involved an official offer of a presidential nomination. Klausner explains that he decided to say good-bye to the NCI and withdraw from consideration for the NIH directorship because he wants to work more directly with scientific research.

"I love this institution, but the reality is that the NIH director does not get closer to the science — he gets further from the science," Klausner says. "And what I was missing more and more was being close to the content."

At the Case Institute, Klausner plans to use his initial \$100-million budget to generate research projects taking a multi-disciplinary approach to science and health problems, with particular emphasis on innovative uses of technology.

Meanwhile, it remains unclear when the health department will select an acting director for the NCI or when a candidate for permanent director will receive the attention of an obviously distracted White House. (Klausner announced his resignation at a meeting of the NCI's advisory board right at the time of the terrorist attacks on New York and Washington, making it difficult even for board members to focus on his departure.)

Normally, an institute's deputy director might expect to be named as acting director until a permanent replacement can be found. The NCI's current deputy director is Alan Rabson, the husband of Ruth Kirschstein, acting director of the NIH. But a senior official at the biomedical research agency said that he was unaware of any regulation preventing Rabson, a veteran NCI administrator, from becoming acting director. ■

Top researchers plan to snub fertility conference

Erica Klarreich, London



Antinori: may generate 'an unruly scrum'.

Several of the world's leading experts on reproductive biology have withdrawn from a forthcoming conference in Monte Carlo, saying it has effectively been taken over by Severino Antinori, the Italian doctor best known for his plans to clone humans.

The conference, set for 11–14 October, was originally planned as the third world congress of the International Association of Private Assisted Reproductive Technology Clinics and Laboratories (A PART), a grouping of institutions that conduct fertility treatments.

But on 9 September, the board of A PART withdrew its support for the meeting, saying that Antinori, director of the International Associated Research Institute for Human Reproduction in Rome, and one of the conference organizers, had taken over financial responsibility for the meeting and excluded its other organizers from decision-making.

A PART had hoped that up to 500 fertility clinicians from around the world would attend the meeting. But by 9 September, 41 of the expected 77 speakers had withdrawn, says Wilfried Feichtinger of the Institute for Sterility Treatment in Vienna, and president of A PART.

Ian Wilmut of the Roslin Institute in Edinburgh, Scotland, who led the team that cloned Dolly the sheep in 1997, is among those who cancelled plans to attend the meeting, says Harry Griffin, an assistant director at the institute.

"We viewed the meeting as a way to put the evidence on the limitations of cloning technology directly to the assisted-reproduction community, and create a measured approach to the media," Griffin says. "But when Professor Antinori is involved in media affairs they usually degenerate into an unruly scrum."

The Italian doctor "has turned it into an Antinori congress", says Peter Brinsden, a board member of A PART and medical director of Bourn Hall Clinic, a private infertility clinic near Cambridge in England. "We realized it was going to become a publicity circus."

Antinori responded by saying that the conference would be successful and conducted at a very high scientific level. ■

Lasker medical awards go to five disease researchers

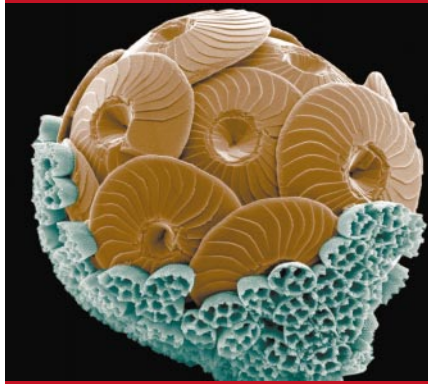
New York The annual Albert Lasker Medical Research Awards — among America's most valued science prizes — are to be presented this week to five scientists in recognition of major advances they have made towards the understanding and treatment of disease.

The award for basic medical research will be shared by Mario Capecchi of the University of Utah, Martin Evans of Cardiff University, UK, and Oliver Smithies of the University of North Carolina for their work on the technology used to create transgenic mice for disease research. Robert Edwards of the University of Cambridge, UK, will receive the clinical research award for his development of *in vitro* fertilization. The award for public service will go to William Foege of Emory University in Atlanta for his role in the eradication of smallpox and the prevention of river blindness and guinea-worm disease.

The awards, which were first presented in 1946, are administered by the Albert and Mary Lasker Foundation, a New York-based charitable organization that funds research into human disease.

♦ <http://www.laskerfoundation.org>

Up close and personal with plankton



The mid-life make-over of a type of plankton has been captured on film with remarkable clarity in this image by Markus Geisen, a micropalaeontologist at London's Natural History Museum. *Calcidiscus leptoporus* changes its appearance during its life, as the old outer covering (green) is replaced with one of plate-like units (brown). The image won Geisen the first prize in the Science Close-Up category of the Novartis/Daily Telegraph Visions of Science Photographic Competition.

♦ <http://www.visions-of-science.co.uk>

Environmentalists batter giant transgenic fish

Munich A new European patent for a technique that creates genetically engineered fish that grow bigger and faster has drawn complaints from environmental groups.

Canadian biotechnology company Seabright, now known as the Genesis Group, the technology-transfer agency of the Memorial University of Newfoundland, filed for the patent in 1992; it was granted in July this year. The patent covers a gene construct for the production of transgenic salmon and

other fish that carry extra genes that speed growth. Atlantic salmon with the gene have growth rates up to eight times higher than controls, the inventors claim.

Environmental groups such as Greenpeace argue that transgenic fish could escape into the wild and become invasive species, damaging fish stocks and marine ecosystems. Transgenic salmon is not currently approved for human consumption. When it is, Seabright hope to sell genetically modified fish eggs to aquaculture farms.

♦ <http://www.genesis.mun.ca>

♦ <http://www.greenpeace.org>

Navy taken to court over sonar test plan

San Diego Environmental groups are suing the US Navy over plans to conduct tests of high-intensity sonar systems off the coast of California. The groups argue that the sonar harms whales, dolphins and other sea life in the area.

Since 1996 the Navy has conducted sonar tests in the Atlantic Ocean, the Gulf of Mexico, the Mediterranean Sea and most recently in seas near Japan. Environmental groups say that the high-power sonar damages the ears of marine mammals, and that whales and dolphins have been found stranded on beaches after the experiments.

The latest tests, which were scheduled for 30 October, are now on hold as the Navy reviews evidence for the environmental impact of the sonar. The lawsuit, filed by the Natural Resources Defense Council and other environmental groups, seeks to have a fuller environmental assessment conducted.

France makes clean sweep in Balzan science prizes

Milan Two French scientists have been awarded the prestigious Balzan prizes for science and culture.

Neurobiologist Jean-Pierre Changeux, of

the Pasteur Institute in Paris, was rewarded for his work on acetylcholine receptors which shed light on how brain cells communicate, and in recognition of his books on philosophy and art. Claude Lorius, director of the CNRS Laboratory for Glaciology and Environmental Geophysics in Grenoble, received the award for his research into how ice can be used as an archive of past climate. Lorius was one of the first scientists to warn of the dangers of global warming.

The International Balzan Foundation, based in Zurich and Milan, was established in 1956 in memory of the Italian journalist Eugenio Balzan, who emigrated to Switzerland in the 1930s after the fascist clampdown on the press in Italy. Each year the foundation awards SFr1 million (US\$625,000) each to two scientists and to two artists, philosophers or writers.

► <http://www.balzan.it>

Art puts the Sherlock Holmes into diagnosis

London Trainee doctors can improve their diagnostic skills by undergoing art-appreciation classes, according to researchers at Yale University in New Haven, Connecticut (see *J. Am. Med. Assoc.* **286**, 1020–1021; 2001).

“Physicians need to be more like Sherlock Holmes,” says Irwin Braverman of the university’s school of medicine. Braverman and colleagues decided to develop students’ sleuthing skills by sending them on a first-year undergraduate class on fine art. After only two hours studying a painting and being questioned on what they saw, the students’ diagnostic skills improved.

Art-appreciation classes are now a part of the curriculum for all Yale medical students. The authors believe that the classes work by accelerating the process of learning to see without preconceived ideas. This is a key part of diagnosis that physicians normally learn over years of clinical practice.



Art attack: paintings such as Henry Wallace's *The Death of Chatterton* could help medics save lives.

Two become one

Developmental biologists and cell biologists have long ploughed their separate furrows. But now these two disciplines are coming together in unexpected and exciting ways, says Helen Pearson.

Kathryn Anderson wasn't looking to embark on an affair with cell biology — it just happened. Anderson, a developmental biologist at the Sloan-Kettering Institute in New York, was studying 'open-brain' mice — which have a fatal genetic mutation that prevents their brain from folding correctly during embryonic development — when her research took a surprising turn.

She had expected the gene involved to be one of the usual suspects implicated in severe developmental abnormalities — those encoding proteins that regulate the expression of other genes, or that are involved in key cellular signalling pathways known to direct embryological development. Instead, she found that the gene produced a protein that carries materials around the cell¹. Such 'housekeeping' proteins had long been ignored by developmental biologists — deemed of interest only to researchers fascinated with the detailed mechanics of a cell.

Anderson is just one of many developmental biologists who are now realizing that apparently mundane aspects of cells' internal workings can help to illuminate some of the mysteries of development. "For whatever reason, developmental geneticists have encountered basic cell biology in the past few years," she says. Cell biologists are likewise coming to recognize that their obsession with the mechanics of processes such as cell division, and the transport of proteins around the cell, can contribute to the study of development. Helping to broker this marriage between fields are techniques of video microscopy, which use fluorescent markers to film proteins moving around in cells and tissues (see 'Lights, camera, action!', overleaf).

For the past two decades, developmental geneticists have been busy identifying 'big hitting' developmental genes. These encode the proteins involved in signalling pathways that direct the organization of animals' tissues and organs. But knowing that the brain, for instance, is directed to grow in the way it does because gene A turns on gene B turns on gene C, and so on, still leaves much to be learned about the developmental processes that underlie the complexity of the finished organ.

"Over the past five years there has been a shift away from the formal logic of genetic pathways," says Stephen Cohen, who studies

the development of wings in the fruitfly *Drosophila* at the European Molecular Biology Laboratory in Heidelberg. This has seen many developmental biologists migrate into the territory staked out by their counterparts in cell biology. In retrospect, says Cohen, it was an obvious move. Developmental biologists consider how tissues and organs form — but these are simply populations of cells working together. So gaining a complete picture of how organisms are built will require an understanding of cells' internal workings.

Steeped in gradients

One area of developmental biology being revitalized by this new wave of thinking is the study of morphogens. These are secreted proteins, many of which trigger the signalling pathways that developmental geneticists have been teasing apart. According to the textbooks, morphogens work their magic by diffusing from a source to form a high-to-low gradient. Cells lying in the path of the diffusing morphogen sense its concentration in order to work out where they sit in a developing organ or tissue and, accordingly, what genes they need to switch on.

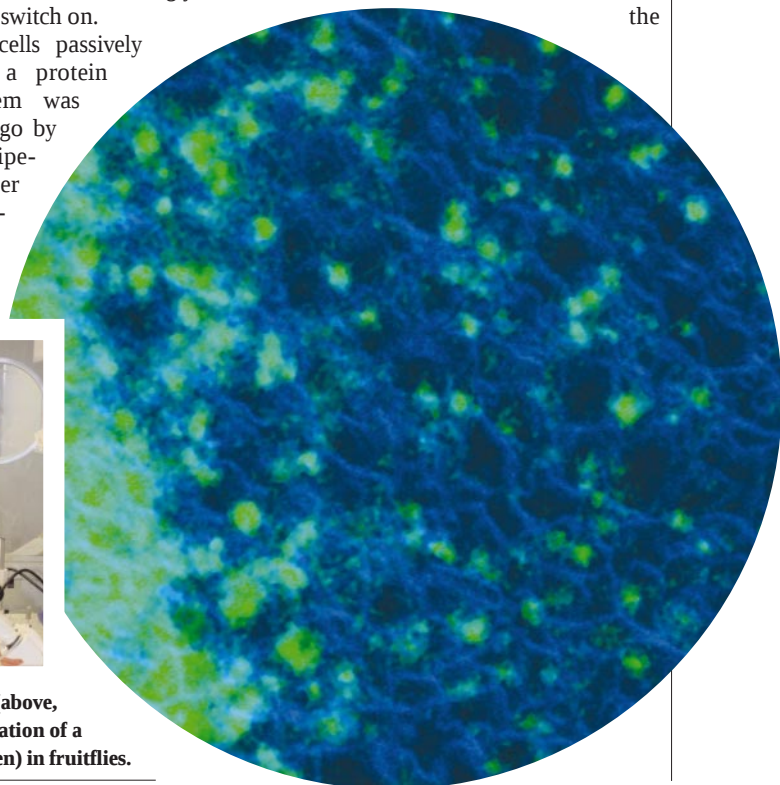
But this view of cells passively receiving cues from a protein diffusing around them was challenged two years ago by Tom Kornberg and Felipe-Andrés Ramírez-Weber of the University of California, San Francisco. Far from just lying down and taking



The mutated gene found in open-brain mouse embryos (right) offers a surprising link between cell biology and development.

their instructions from morphogens, it seems that cells often play an active role in determining how the gradient forms.

Kornberg and Ramírez-Weber were looking at the disc of developing tissue that ultimately gives rise to a fruitfly wing. They labelled cells at the edge of the disc with green fluorescent protein (GFP), a commonly used fluorescent marker. This revealed long thread-like fingers, which the researchers dubbed 'cytonemes', extending from these cells towards the



Marcos González-Gaitán (above, rear) has imaged the formation of a morphogen gradient (green) in fruitflies.

disc's dark, unlabelled centre, where a morphogen called Decapentaplegic (Dpp) is produced². Cells actively seek out morphogens by extending these protrusions, Kornberg and Ramírez-Weber proposed — although they have yet to prove that morphogens are actively channelled along the cytonemes.

Disc drive

Other researchers have labelled key fruitfly morphogens such as Hedgehog, Wingless and Dpp with fluorescent markers and watched the concentration gradients form. "It has changed the way I look at things," says Marcos González-Gaitán of the newly opened Max Planck Institute of Molecular Cell Biology and Genetics in Dresden (see this issue's *NatureJobs*, pages 4–5), which is concentrating on the interface between cell biology and development.

By fusing Dpp to GFP, González-Gaitán and his colleagues were able to watch a gradient form in the fruitfly wing disc. Some of the Dpp could clearly be seen in small round dots inside cells that the researchers think are endosomes — membrane-bound vesicles that form when substances are actively taken up by cells. And in discs containing patches of mutant cells unable to form endosomes, the flow of the morphogen through the tissue was impeded³. "Instead of going between cells it's going through them," says González-Gaitán.

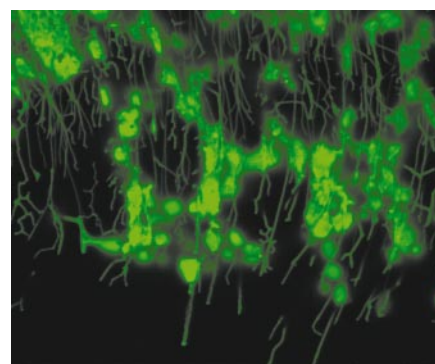
This means that the uptake and subsequent release of morphogens by individual cells may help to determine how quickly and how far they move through a tissue, and so determine the 'shape' of the concentration gradient that forms.

A related cellular process — the speed at which morphogens are broken down after being taken up into endosomes — may also be key in shaping morphogen concentration gradients. Wingless is a morphogen that normally accumulates in a series of graded stripes across the body of a developing fruitfly embryo, and is involved in specifying the insect's segmented body plan. But when Jean-Paul Vincent of the National Institute for Medical Research in London blocked the protein's breakdown using a drug called chloroquine, the concentration gradient changed, with the Wingless stripes extending into areas that do not normally contain the morphogen⁴.

This month, Suzanne Eaton, also at the Dresden Max Planck Institute, published another observation that may help to explain morphogen trafficking. She was studying how proteins are sorted to one side or the other of the fruitfly wing disc. She tagged GFP to a membrane protein expressed on only one side, but she noticed tiny glowing dots inside cells on the other, unlabelled side. These dots seem to form as tiny membrane-bound packages within larger endosomes. Eaton calls them 'argosomes'⁵, and speculates that they are released when endosomes fuse with cells' outer membranes — allowing morphogens to be passed from one cell to another.

The mechanisms being touted to explain the formation of morphogen gradients are not mutually exclusive, and it is possible that different processes come into play for different morphogens, and for the movement of the proteins over different distances. "Everyone would like to be the one who figures it out," says Eaton.

Other developmental biologists, meanwhile, are moving into detailed studies of cell division. It has long been known that asymmetric cell division, in which the two daughter cells that arise from an unequal cell division acquire different fates, is a fundamental way of creating cell diversity during development.



Green-fingered: are the thread-like cytonemes that reach out from cells searching for morphogens?

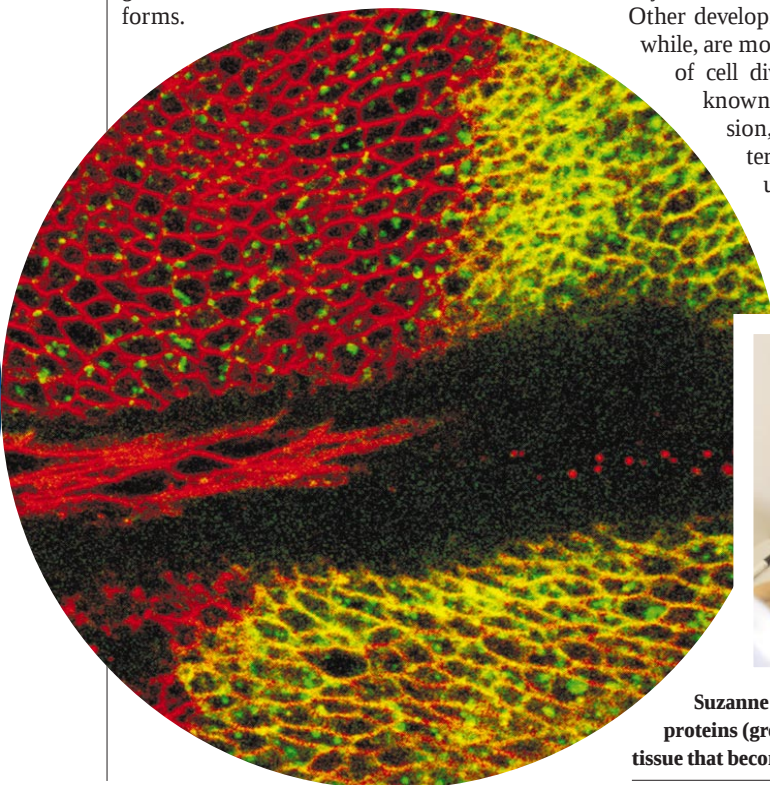
For this to achieve the desired results, the cells must be 'polarized' — they must have a 'top' and a 'bottom' in relation to the developing body plan. The fate of the daughter cells depends on the distribution of key proteins called determinants. Over the past few years, researchers have shown that asymmetric cell division depends both on the localization of these proteins within the parent cell before division, and on the orientation of a structure called the mitotic spindle, along which replicated chromosomes move to opposite ends of the cell before it divides. The spindle determines the plane of cell division, and its positioning influences the relative sizes of the resulting daughter cells, and the quantities of determinants that each receives.

Determined efforts

Back in 1994, a protein determinant called Numb was shown to accumulate in a crescent at one end of cells in fruitfly embryos that give rise to neurons⁶. Since then, researchers have identified a host of other determinants that also become localized to one or the other end of asymmetrically dividing cells, and they are busy investigating how this localization occurs⁷.

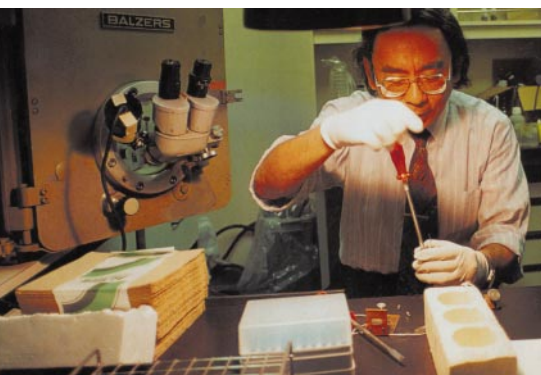
Meanwhile, other groups are investigating the orientation of the mitotic spindle. Andrea Brand and her team at the University of Cambridge, for instance, have revealed that a sudden shift in the spindle's axis is crucial to the development of the fruitfly nervous system. The cells that give rise to neurons develop from a simple layer of cells in which Numb, and another determinant called Prospero, are concentrated at the base of each cell. In the divisions that form this layer, the plane of cell division runs from top to bottom of the parent cell — which means that equal quantities of the determinants end up in the daughter cells. But some cells, called neuroblasts, detach from the layer and divide asymmetrically into a large neural stem cell and a smaller cell that inherits Prospero and Numb and subsequently divides to give rise to two neurons.

Brand's team showed that the change between symmetric and asymmetric division involves a simple switch of spindle

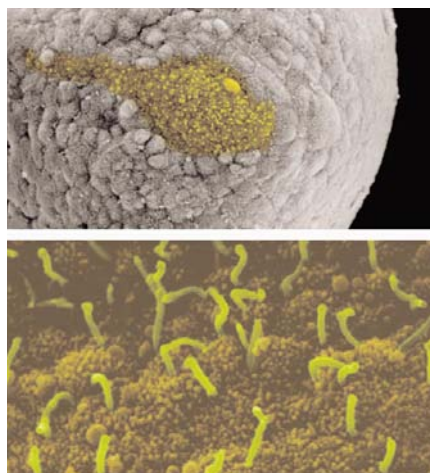


Suzanne Eaton saw membrane proteins (green) migrate across the tissue that becomes a fruitfly's wing.





Cilia (right) on cells in the 'node' (top right) may direct the asymmetrical development of important organs, says Nobutaka Hirokawa.



alignment. The researchers fused GFP to tau, a protein component of the mitotic spindle, and tracked the spindles using time-lapse video microscopy. In neuroblasts, they found that the spindle rotates by 90° in 60 seconds just before an asymmetric division⁸. "When I show the movies there are often gasps from the audience," says Brand.

Another key issue is understanding how the spindle positions itself asymmetrically in cells that divide unequally. Earlier this year, a team led by Tony Hyman of the Dresden Max Planck Institute shed light on the process by studying the first cell division in embryos of the nematode worm *Caenorhabditis elegans*. In this species, early asymmetric cell divisions determine cell fates for the entire organism.

Hyman's team severed the mitotic spindle with an ultraviolet laser and used time-lapse video microscopy to watch what happened to structures called centrosomes, to which the two ends of the spindle are attached. Both began to move off towards the ends of the cell, but one moved farther and more rapidly than the other⁹. This indicated

that the asymmetric positioning of the spindle occurs as a result of differential pulling forces exerted on the centrosomes by microtubules in the cell's internal skeleton. By studying mutant worms, the researchers also revealed that these forces are controlled by the genes that determine cell polarity.

Two-way traffic

Although many of the players in developmental cell biology are migrating from traditional developmental biology, they are being joined by cell biologists moving in the opposite direction. For instance, until three years ago, Nobutaka Hirokawa of the University of Tokyo was tightly focused on the proteins that transport cargoes of other proteins around the cell.

But when Hirokawa and his colleagues engineered mice to lack one such protein, called KIF3B, they encountered a developmental quirk. The mice died during gestation — but in half of them, the heart, liver and other organs had begun to form on the opposite side of the body to their usual

position¹⁰. "Until this moment we were not developmental biologists," says Hirokawa. But he knew that the formation of body-plan asymmetries was a hot topic in developmental circles, so he pitched in.

Hirokawa and his colleagues took a fresh look at the node, a small structure formed at the end of what eventually becomes the spinal cord — and which is thought to be involved in establishing left-right asymmetry. Cells in the node each have hair-like cilia protruding from their surface. Hirokawa's team revealed that these rotate at 600 revolutions per minute. Fluorescent beads thrown into the surrounding fluid were tugged from one side to another by the current generated.

In mice lacking KIF3B, the cilia do not form, because the protein transports components needed for their construction. Hirokawa suggests that the rotating cilia set up a morphogen gradient that directs the normal pattern of asymmetric gene activity and causes organs to form on the correct side of the body. If the cilia are absent, he argues, this asymmetry develops at random. But many developmental biologists remain unconvinced because the putative morphogen remains unknown. "It's provocative," admits Hirokawa. "Cell biologists like that."

Despite occasional differences in perspective, the distinctions between cell biologists who are becoming interested in development and developmental biologists who are discovering cell biology are becoming blurred. "Merging is inevitable," says Andy McMahon, who studies mouse development at Harvard University. He believes the biggest potential impediments are technical, such as the difficulty of transferring cell-biology techniques into more complex systems. Although it has proved easy, for instance, to transfer cellular imaging techniques to fruitfly embryos, mouse and chick embryos are more bulky, making it difficult to visualize individual cells within a developing tissue or organ.

To get round this problem, groups such as McMahon's are establishing methods to grow individual organs or tissues in culture. If they succeed, developmental and cell biology should be poised for a long and happy marriage. Hyman says that for years he endured the scorn of developmental colleagues who could not understand his interest in cell biology. "Not any more," he says. ■

Helen Pearson works in Nature's science writing team.

1. Eggenschwiler, J. T., Espinoza, E. & Anderson, K. V. *Nature* **412**, 194–198 (2001).
2. Ramírez-Weber, F. A. & Kornberg, T. B. *Cell* **97**, 599–607 (1999).
3. Entchev, E. V., Schwabedissen, A. & González-Gaitán, M. *Cell* **103**, 981–991 (2000).
4. Dubois, L., Lecourtis, M., Alexandre, C., Hirst, E. & Vincent, J.-P. *Cell* **105**, 613–624 (2001).
5. Greco, V., Hannus, M. & Eaton, S. *Cell* **106**, 633–645 (2001).
6. Rhyu, M. S., Jan, L. Y. & Jan, Y. N. *Cell* **76**, 477–491 (1994).
7. Knoblich, J. A. *Nature Rev. Mol. Cell Biol.* **2**, 11–20 (2001).
8. Kaltschmidt, J. A., Davidson, C. M., Brown, N. H. & Brand, A. H. *Nature Cell Biol.* **2**, 7–12 (2000).
9. Grill, S. W., Gönczy, P., Stelzer, E. H. & Hyman, A. A. *Nature* **409**, 630–633 (2001).
10. Nonaka, S. et al. *Cell* **95**, 829–837 (1998).

Lights, camera, action!

Attend any conference on developmental cell biology, and you'll spend some of your time watching videos. "Everyone wants to make movies," says Bruce Bowerman, who works on the mechanics of cell division at the University of Oregon in Eugene. "It's like Hollywood."

Indeed, the field's emergence has been facilitated by technology to make videos that reveal how development is unfolding at the cellular level in living tissues.

The key to this technology was the isolation of green fluorescent protein (GFP) in the jellyfish *Aequorea victoria*. Using genetic engineering, GFP can be fused to various cellular components without disrupting their normal function. And biologists bored with green can now extend their palette. GFP has been genetically tweaked into shades of yellow, blue and cyan — and a red version has been isolated from *Discosoma* mushroom anemones, a type of soft coral.

Tracking the movement of fluorescent proteins involves time-lapse video imaging of living cells or tissue slices using sophisticated cameras and microscopes (see left). Lasers are used to excite the fluorescent dyes, cutting down on background fluorescence and allowing researchers to focus on thin slices within a larger sample of tissue. Images of adjacent slices can also be stacked up together to give a three-dimensional picture.



The stowaways

Over the coming decade, exploration of Mars may reveal whether or not life ever existed on the red planet — but only if the missions can avoid detecting any microbes they bring with them, says Tom Clarke.

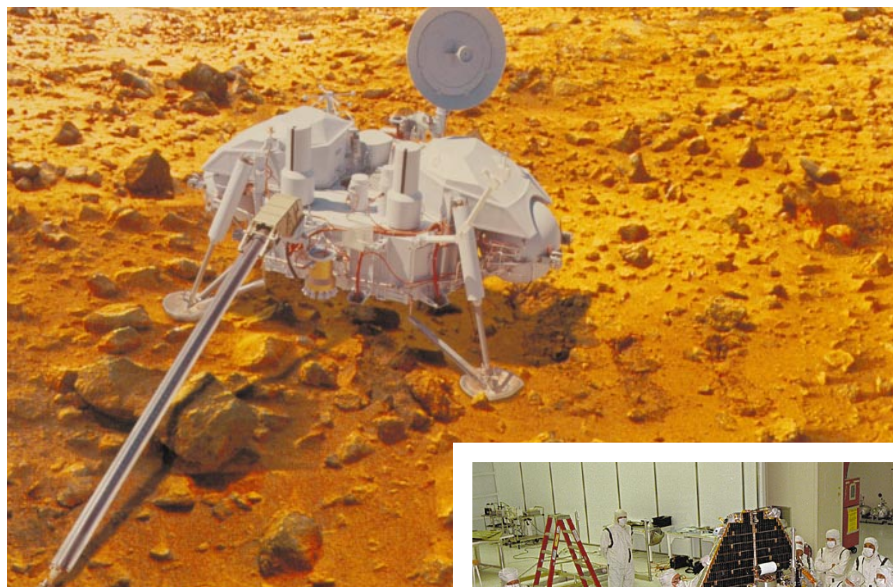
In 1976, NASA's two Viking landers bumped down on the surface of Mars and began their search for life. They failed to find any, showing instead that the martian topsoil is extremely inhospitable — highly oxidizing and continually bombarded by massive doses of ultraviolet radiation.

After Viking, most scientists believed Mars to be lifeless. But the subsequent discovery of bacteria living in extreme Earthly environments, such as in Antarctic ice and the near-boiling, sulphurous waters of deep-sea hydrothermal vents, together with mounting evidence that water once flowed over the red planet, has rekindled hope that there could be — or at least could once have been — life on Mars.

Now NASA and the European Space Agency (ESA) are set to embark on a series of new missions to Mars. But a lingering problem remains — how to ensure that the landers' highly sensitive life-detection instruments are not fooled by biological molecules that have hitched a ride from Earth. Kenneth Nealson, an astrobiologist at the California Institute of Technology in Pasadena, says the field is haunted by the possibility of a humiliating embarrassment that it would never live down: "Thinking you've detected life on Mars and finding that it's *Escherichia coli* from Pasadena."

Before Viking, fears that terrestrial life could contaminate other planets meant that techniques for sterilizing spacecraft were given high priority. But once Viking showed that Mars is almost certainly incapable of supporting life from Earth, the issue slipped down the agenda. "Current technologies are basically the same as they were in the sixties and seventies," says Bob Koukol of NASA's planetary protection group, based at the Jet Propulsion Laboratory in Pasadena.

Worried about the potential for false positives in their search for martian life, both NASA and ESA are experimenting with a range of new anti-contamination techniques. Neither agency has time to spare. ESA's Mars Express mission includes Beagle 2, a British robot that will sample the martian surface and atmosphere.



Sterilizing technology in the facilities where spacecraft are assembled (right) has changed little since Viking's mission to Mars in the 1970s.

Mars Express should launch in mid-2003 and arrive later the same year. NASA plans to land a clutch of rovers on the planet in 2004, followed by further missions in 2007.

Viking invaders

Researchers involved in the search for new methods to sterilize spacecraft face several difficulties. Traditional techniques — the Viking probes were sterilized by heating to around 100 °C for 30 hours — are not applicable to modern spacecraft with their fragile microprocessors and composite materials. And although spores of at least one type of bacterium were known to have survived Viking's heat treatment, others have since been added to the list.

Even if all the microorganisms contaminating a spacecraft are killed, it may not be enough to avoid a false-positive result. Modern life-detection equipment can spot tiny amounts of carbon-containing compounds — enough for it to be fooled by dead bacteria left behind after cleaning and sterilization, or even the organic chemicals used in cleaning. "This is the biggest challenge," says John Rummel, planetary protection officer at NASA's headquarters in Washington.

The NASA and Beagle 2 teams are reluctant to discuss the details of the methods they are working on, as each is keen to commer-

cialize any techniques they develop. Both groups are also wary of potential criticisms of their methods, and say they will not publish them until they are finalized.

But the teams confirm that they are interested in a technique used to disinfect surgical equipment in hospitals. Devices are sterilized in a vacuum chamber into which a sample of hydrogen peroxide is injected. This vaporizes, and a high-voltage electric field is then used to strip the electrons from the gaseous atoms, leaving behind a plasma containing highly reactive atomic nuclei. These react with and destroy any biological molecules on the surface of the device, without unduly damaging the equipment itself.

Mark Sims, Beagle 2's project manager at the University of Leicester, says his team will definitely be using plasma sterilization on some of Beagle's metal parts, but other areas will have to be treated in different ways. Parts of the craft that come into direct contact with samples may also be heat-treated, he says.

The teams are also testing alternative ways of cleaning the probes before sterilization. The traditional technique — wiping with



John Rummel: cleaning is the big challenge.

▶ alcohol — can compound the problem. On aluminium, for example, this bursts cells and glues their innards to the surface.

NASA is currently experimenting with alternatives, including ultrapure water and the water-based cleaning fluids that the semiconductor industry is developing to replace the chlorofluorocarbons used to clean circuit boards. So far, no single method has proved adequate for aluminium — a potential problem as the metal is a common component of spacecraft.

The nylon brushes used to scrub surfaces before wiping with alcohol can also cause difficulties. To a mass spectrometer — standard equipment for detecting the molecular signatures of life — the bonds between nitrogen atoms in nylon could easily be confused with those in amines, the building blocks of proteins. NASA has now taken the advice of David White, an environmental microbiologist at Oak Ridge National Laboratory in Tennessee, who suggested using polypropylene instead, which should not confuse the analysis.

The teams are also working on techniques

to determine how well their attempts at cleaning and sterilization have worked. Because removing all traces of contaminating microorganisms is impossible, it is important to estimate the number of microbes that remain. Upper limits agreed before Viking reflect the level of cleanliness that was achievable at the time — in this case, a maximum of 300 bacterial spores per square metre on any one part of the probe's surface, together with an upper limit of 300,000 spores for the whole craft.

Glowing reports

NASA is currently adapting standard microbiology techniques to assess the precise level of contamination. Amoebocytes from the horseshoe crab — the equivalent of our white blood cells — cluster around dead and living membranes of some types of bacterium. Unlike the membranes themselves, these clusters can be detected by standard laboratory techniques. So researchers can estimate the number of membranes, and hence the number of dead and living microbes, by measuring the number of clusters.

Luciferase, the enzyme that gives fireflies their glow, is also being used. The enzyme is powered by adenosine triphosphate (ATP), which is found in all living or recently living cells. Samples from the surfaces of spacecraft are mixed with luciferase that has been starved of ATP. The level of light produced can then be used to infer how much ATP is in the sample, which in turn indicates the amount of contamination present. But it is a crude test, Rummel admits, as it cannot distinguish between a few large cells, such as yeast, and a greater number of smaller bacterial cells.

Infrared spectroscopy provides the ultimate assessment of contamination, as it can in

principle detect individual molecules by analysing the way they interact with light. Spectroscopy can only be used to study tiny samples, but the Beagle 2 team plans to use the technique to assess contamination on the parts of the craft that will come into direct contact with martian samples.

The NASA and Beagle 2 teams will also monitor contamination in their spacecraft-assembly facilities, in order to provide a list of potential microbial stowaways. If they do end up detecting a possible martian microbe, the inventory will act as a background against which the signal can be compared. "We definitely want to be able to recognize terrestrial life that has hitched a ride there," says Max Coleman, a geologist at the University of Reading in England who is advising NASA on its detection techniques.

Bugs in the system

Compiling this inventory is "an awful job", says White. Numerous bacteria in the facilities cannot be cultured and identified, a problem NASA is hoping to solve by looking for traces of RNA from their ribosomes, cellular structures involved in the synthesis of proteins. Researchers have identified ribosomal RNA sequences for many types of bacterium which have never been cultured. The technique could provide the most thorough inventory of potential bacterial stowaways, but it is proving difficult to implement. "The bugs haven't been worked out yet," concedes Rummel. The British team behind Beagle 2 is likely to face similar difficulties when it starts assembling its spacecraft next July.

When the full protocols are finalized, Rummel says that NASA will go public with its methods. But this may not happen before the 2007 missions — too late for the Beagle 2 team to benefit from them. "We have tried to pick their brains and have had some discussions," says Sims. "But like us, they are developing new procedures." Full details of the Beagle 2 proposals are contained within a planetary protection plan that has just been submitted to ESA, and which the Beagle 2 team insists must remain confidential.

Some experts, such as White, argue that the secrecy surrounding the two teams' work could threaten the reliability of the life-detection experiments. Ultimately, the intense desire to avoid a false positive may provide the motivation for all those involved to begin discussing their work more openly. As Coleman says: "The possibility of subsequent embarrassment is too great."

Tom Clarke is a member of Nature's science writing team.

♦ <http://www.beagle2.com>

♦ <http://planpro.jpl.nasa.gov>



David White: advising NASA on microbial contamination.



Life detectors: cells from the horseshoe crab (below) and an enzyme from fireflies may provide ways to identify terrestrial molecules.



G. SCOTT/SPL

J.L. ROTMAN/CORBIS

Probabilities will help us plan for climate change

Without estimates, engineers and planners will have to delay decisions or take a gamble.

Sir—In his Commentary “What is ‘dangerous’ climate change?”, Stephen Schneider¹ argued that—in the absence of unambiguous expert advice—decision-makers will produce their own probability estimates about future climate change within the large range of projections provided by the Intergovernmental Panel on Climate Change (IPCC) in its Third Assessment Report, and that this is worse than using informed estimates provided by relevant experts. On the other hand, Grübler and Nakicenovic² in a subsequent Correspondence, “Identifying dangers in an uncertain climate”, argue that providing well-founded probability estimates is difficult or impossible because, although probabilities in the natural sciences are based on repeated experiments and frequencies of measured outcomes, this is not the case in the socio-economic sciences.

However, this frequentist basis for probabilities in predictions of an unknown future is not possible in the Earth sciences either, since there will only be one real outcome, which cannot be measured now. Probability estimates of future conditions on Earth based on modelling are not frequentist, but essentially bayesian³, in that they are based on prior knowledge or assumptions embodied in the various models and inputs.

Various authors have shown that such estimates for global warming are possible^{1,4,5} by deriving single-peaked probability distributions. On the other hand, Gritsevskiy and Nakicenovic⁶ obtain a bimodal probability distribution of carbon dioxide emission amounts for the year 2100, which they attribute to a bifurcation in technological development pathways. If this bimodal distribution were combined with other factors to derive a probability distribution for global warming, it would almost certainly be smoother, but could still differ significantly from the other distributions.

Thus, although probability estimates are needed, methods for deriving probabilities require further development⁷. Some authors have suggested that in the meantime we should rely on increasing robustness⁸ or resilience⁹ in developing mitigation and adaptation responses. But without well-founded probability estimates, these strategies still tacitly assume some estimates of likelihood to limit the magnitude of their responses in relation to costs.

In fact, a risk-management approach requires not an assessment of the probability of a particular amount of

greenhouse-gas emission or global warming at some future time, but rather an estimate of the likelihood of exceeding an identified critical impact threshold¹⁰. This integrates the probabilities from the least climate change up to the critical level, and is much more robust with respect to the underlying assumptions.

In the IPCC Third Assessment Report discussed in ref. 1, the uncertainties in projections of global surface warming by 2100 derive almost equally from uncertain emissions and uncertain climate science. Thus if critical thresholds for impacts lie in the top half of the predicted range of global warmings, these are much more likely to be avoided if decision-makers attempt to follow socio-economic development pathways that will reduce emissions. The virtue of the range of socio-economic development pathways explored by IPCC in ref. 11 is that it suggests that there are various plausible socio-economic futures that will achieve this goal. Such development pathways may still lead to appreciable warming by 2100 and beyond, which can be avoided only by stringent climate policies.

Without probability estimates,

engineers and planners will be left needing to foster resilience and adaptive capacity, hedge their bets, delay their decisions, or else gamble on whether humanity will go down high or low emissions development pathways as they adapt design standards and zoning to climate change.

It is far more sensible to establish cumulative probability distributions to allow optimal, focused adaptation plans.

A. Barrie Pittock, Roger N. Jones, Chris D. Mitchell

Climate Impact Group, CSIRO Atmospheric Research, PMB 1, Aspendale 3195, Australia

1. Schneider, S. H. *Nature* **411**, 17–19 (2001).
2. Grübler, A. & Nakicenovic, N. *Nature* **412**, 15 (2001).
3. Malakoff, D. *Science* **286**, 1460–1464 (1999).
4. Webster, M. D. et al. *Uncertainty Analysis of Global Climate Change Projections*, Report No. 73, MIT Joint Program on the Science and Policy of Global Change (Cambridge, MA, 2001).
5. Wigley, T. M. L. & Raper, S. C. B. *Science* **293**, 451–454 (2001).
6. Gritsevskiy, A. & Nakicenovic, N. *Energy Policy* **28**, 907–921 (2000).
7. New, M. & Hulme, M. *Integrated Assessments* **1**, 203–213 (2000).
8. Lempert, R. J. & Schlesinger, M. E. *Climatic Change* **45**, 387–401 (2000).
9. Barnett, J. *World Development* **29**, 977–993 (2001).
10. Jones, R. N., *Natural Hazards* **23**, 197–230 (2001).
11. Nakicenovic, N. & Swart, R. *Special Report of the Intergovernmental Panel on Climate Change on Emissions Scenarios* (Cambridge Univ. Press, Cambridge, 2000).

Vital parameters need to be in print

Sir—In an effort to condense Letters to *Nature* in the printed version of the journal, there is a risk that some of the critical information necessary for an independent judgement of the quality of biological structural information may be omitted, appearing only in the Supplementary Information available in the online version.

These indicators are: the resolution of the structure determination; the ‘free R factor’ (an unbiased metric of agreement between the experimental X-ray diffraction data and the derived molecular model); and the Ramachandran analysis (the only independent measure of a model’s stereochemical reasonableness).

In some cases, this critical information—which takes little space—is excluded, whereas non-essential information, such as the location of the synchrotron beamlines where the diffraction data were collected and the names of standard programs used to determine the structure, is included in the printed version. It is often inconvenient to have to access the Internet while reading a paper in print, in order to obtain these essential quality indicators.

The crystallographic community has had a welcome change of heart on agreement about validation criteria for structural models—critical use of the indicators I have mentioned above—and on data availability via deposition in publicly available databases. Given that *Nature* has fully supported this shift, should the journal not also enforce minimal, accessible reporting of a structure’s quality indicators in the printed versions of papers, rather than allowing authors to deposit all this information in the online-only format of Supplementary Information?

David Borhani

Abbott Bioresearch Center, 100 Research Drive, Worcester, Massachusetts 01605, USA

Nature’s policy is to include in the printed versions of its papers sufficient technical information to allow an interested reader to appreciate the experiments that have been performed, along with the most critical experimental data. We would therefore normally expect essential structural parameters to be included in the published version of papers rather than as part of the Supplementary Information only. However, this is always assessed by *Nature* on a case-by-case basis—Editor, *Nature*.

Paradox of the savant mind

The provocative exceptions to our understanding of intellectual ability.

**Bright Splinters of the Mind:
A Personal Story of Research
with Autistic Savants**

by Beate Hermelin

Jessica Kingsley: 2001. 160 pp. £29.95, \$49.95
(hbk); £13.95, \$19.95 (pbk)

Allan Snyder

Art, music and mathematics are often presumed to be the supreme expressions of the creative mind. They supposedly require superior intelligence along with years of disciplined training. But do they? By the age of three, Nadia could draw horses with astonishingly lifelike perspective. She did so spontaneously, without any training, and from memory. Yet Nadia could not even distinguish her mother from the nurse and was unable to communicate.

Tom, from the age of four, could play Mozart piano sonatas flawlessly after a single hearing. He could also repeat, word for word, extended conversations in any language. Yet he, too, was mentally retarded and lacked the ability to communicate. The same was true of Joseph, the inspiration for the Hollywood film *Rain Man*, who could mentally calculate which two numbers, when multiplied, give the product 1,234,567,890.

These provocative exceptions to our understanding of intellectual abilities are known as savants. Savants are extremely rare individuals who, although severely brain damaged, display islands of astonishing excellence. And they have no idea how they do it.

How do they do it? One view argues that savants acquire their skills like any normal person, through repetitive practice. The idea is that brain damage fuels obsessive interest and the capacity for pathological concentration. But this explanation does not fit well with the fact that savant skills often emerge 'spontaneously' and do not improve qualitatively with time, even though the skill might become better articulated.

At the other extreme, my colleagues and I have argued that savant skills are a form of mimicry, requiring little or no practice. Savants simply tap into information and mental machinery that resides equally in us all but cannot normally be accessed. For example, our brains possess algorithms for calculating the shape of an object from subtle shading across its surface. We are not conscious of this shading, for otherwise we would all be able to draw without training. But brain damage enables savants to have privileged access to such information.

To formulate explanations such as these, we need to have quantitative studies of savant



Nadia's horses, drawn when she was aged three (inset) and five. But Nadia couldn't tell her mother from her nurse and was unable to communicate.

performance. Most accounts of savants are purely descriptive. But in the 1960s, two trail-blazing psychologists, Beate Hermelin and Neil O'Connor, brought autistic savants into the laboratory and began to devise methodologies and conduct carefully controlled experiments to unravel this phenomenon.

Bright Splinters of the Mind is Beate Hermelin's personal account of her research with autistic savants, in particular their abilities in poetry, foreign languages, lightning-fast arithmetic, drawing and calendar calculating. Discussing artistic savants, Hermelin concludes that they have a "specific ability to

draw", and that they use the same pictorial rules of linear perspective as do trained normal artists to portray a three-dimensional world on a flat surface.

Take mathematical savants who can rapidly identify whether or not a large number is a prime (divisible only by one or by itself). Hermelin concludes that these savants use the same "strategies as mathematically trained individuals", specifically the Eratosthenes algorithm — derived from the 'sieve' of the Alexandrian philosopher Eratosthenes for identifying prime numbers. A similar conclusion is reached for calendar-calculating savants who can instantly answer

FROM NADIA: A CASE OF EXTRAORDINARY DRAWING ABILITY IN AN AUTISTIC CHILD BY L. SELFE (ACADEMIC, 1977)

questions such as "what day of the week was 18 April 1720?". Hermelin says that they use the rules and regularities of the calendar, and that musical savants extract the "grammar" of music.

So Hermelin believes that savants apply the same rule-based strategies as do trained people of normal intelligence. How do they learn these strategies? She believes that the rules of linear perspective used in drawing are extracted from posters and illustrations. As for the other skills, she believes savants advance from a focus on specific details (say, numbers) to the whole picture (say, the Eratosthenes algorithm).

But savant skills can emerge suddenly after a person is hit on the head, so it seems possible that these skills are in us all without training, but cannot normally be accessed. Recent evidence suggests that they might even be switched on by using magnetic pulses to switch off part of the brain, as our work had indicated.

Hermelin's is a highly readable book. She goes well beyond merely presenting a scientific account. Rather, she conveys something about who these people really are. She weaves a tapestry of their personal lives, especially their difficulties in confronting life as we normally know it. The book works well at all levels.

Anyone who has interacted with autistic individuals will appreciate the magnitude of Hermelin's contribution. Her findings are a giant step forward in unravelling the

treasures of our minds. To extract the core explanation for savant skills, it might be necessary to test savant prodigies when their skill first emerges because, with maturity, autistic savants often acquire concepts and knowledge which inevitably become incorporated into their skill base. Such research remains a herculean task for future investigators. ■

Allan Snyder is at the Centre for the Mind, Australian National University, Canberra, ACT 0200, and University of Sydney, Main Quadrangle, Sydney, New South Wales 2006, Australia.

Spandrels or selection?

The Evolutionists: The Struggle for Darwin's Soul
by Richard Morris
W. H. Freeman: 2001. 272 pp. \$22.95, £18.99

Dawkins vs. Gould: Survival of the Fittest
by Kim Sterelny
Icon: 2001. 160 pp. £5.99, \$9.95 (pbk)
Michael A. Goldman

Nature or nurture? Chance or necessity? These dichotomies embody a controversy that has raged among the top thinkers in evolutionary biology. The question is: does adaptation by natural selection explain everything in nature, including human

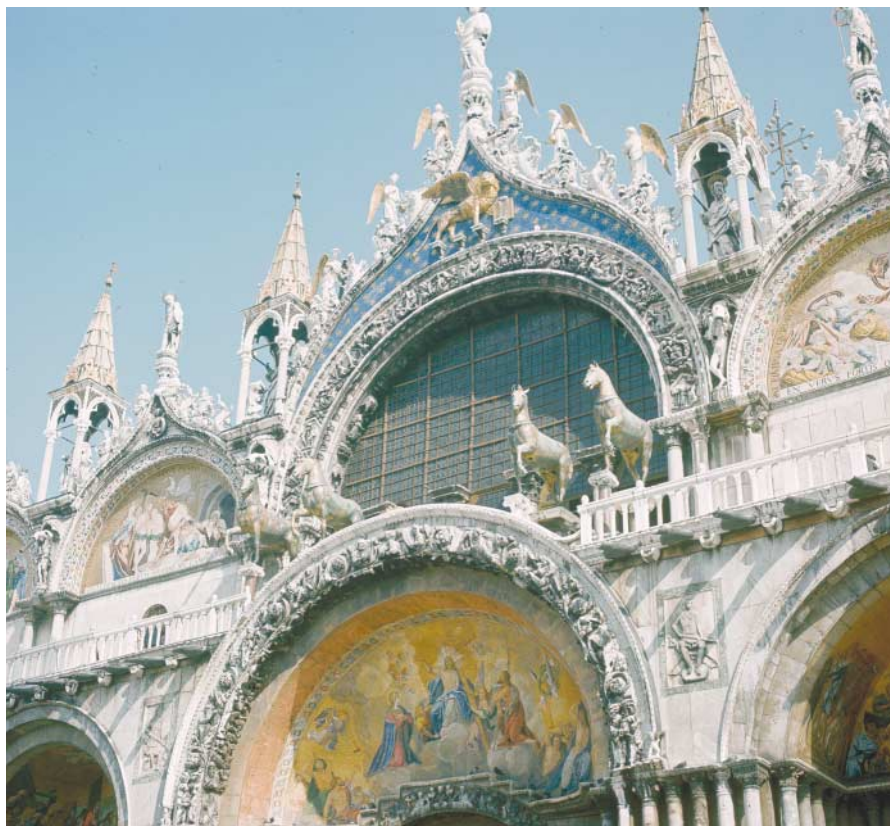
behaviour, or is the situation more complicated? The problem is that no one really believes the first proposition, but the second does not constitute a useful scientific hypothesis. And except as the impetus for a spate of books and articles, and lots of acrimonious debate, it may not matter much.

The contemporary debate started in 1979, when Stephen Jay Gould and Richard C. Lewontin published an article entitled "The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme". This became the focus for the conflict between two lines of evolutionary thought. On one side are Richard Dawkins and like-minded evolutionary biologists, who believe that natural selection is adequate to explain virtually every observation in evolutionary biology. On the other are Gould and his followers, who believe that natural selection is a very important force in evolution, but not the only one. The most heated controversy arises when we attempt to apply our knowledge of evolutionary biology to the origin of human behaviour.

In *The Evolutionists* and *Dawkins vs. Gould*, Richard Morris and Kim Sterelny, respectively, recount this controversy in excruciating detail. Sterelny gets almost to the heart of the matter, and Morris's engaging style makes the history, politics and political motivations fun to read. Unfortunately, neither author really brings us any closer to a resolution, and neither really explains why the controversy may never be resolved.

Both try to dissect the argument into its component parts. They agree that Gould departs from "Darwinian fundamentalists" in his belief that evolution occurs by periods of stasis followed by periods of rapid evolution ("punctuated equilibria" or, as his detractors quip, a theory of "evolution by jerks"). Palaeontologist Gould sees evidence for rapid transitions, catastrophic extinctions and spectacular radiations in the fossil record, and thinks that a model of slow, steady change by natural selection acting on genetic variation is not adequate to explain history. In particular, Gould's notion of contingency in evolution may be important in understanding the origin of new species and higher taxa, and aspects of the broad pattern of evolutionary history that have never been fully explained by the neodarwinian synthesis.

Another area of disagreement concerns Gould and Lewontin's concept of 'spandrels' in evolution. Named after an architectural feature that is a by-product of the construction, evolutionary spandrels are biological structures or traits that are accidental by-products of history, not the results of natural selection. However, natural selection can clearly mould a spandrel into a useful structure. Spandrels, Morris and Sterelny agree, don't much change our understand-



The spandrels of San Marco's basilica, symbols of a long-running debate on evolution.

ing of anatomical evolution. But the issue becomes very heated where sociobiology or evolutionary psychology are concerned. Gould believes that many human behavioural traits are spandrels — by-products of the brain we evolved in the African savannah; the ability to read *Nature* is a spandrel, not a product of natural selection. The Dawkins party tends to think of the brain as a collection of traits moulded by natural selection. Morris gives an elaborate recipe, and even some preliminary data, for deciding between these two views by examining whether the brain is composed of isolated functions or parts or is an interacting whole. But anyone who thinks a brain is composed of interacting parts, whereas a body is not, has never suffered a stiff neck as a result of limping with a sprained ankle, yet no one is arguing that ankles and necks aren't largely products of natural selection.

Morris devotes a chapter to complexity theory, providing a lucid and enlightening explanation. Complexity scientist Stuart Kauffmann "believes that, although natural selection is important, it is not the sole cause of evolutionary change", which is a "marriage between self-organization and selection". Complexity science indicates that there are "emergent properties" that could not be predicted by a reductionist approach, a view that pleases Gould.

Why is it so important that we know whether human behavioural traits are spandrels, and whether human brains have emergent properties? The answer lies more in politics and philosophy than it does in science. Sterelny explains most of the conflict between Dawkins and Gould in terms of two distinct ideologies. "In short," Sterelny contends, "Dawkins, but not Gould, thinks of science as a unique standard-bearer of enlightenment and rationality." Dawkins views the entire world in reductionist terms, and is dedicated to the scientific method as the only valid mode of analysis. Gould, as he has written elsewhere, sees science and religion as "non-overlapping magisteria", as spheres in which different sorts of reasoning apply. He is probably right. But he also views some of science itself as outside the realm of investigation. Dawkins thinks that modern evolutionary theory provides a good model for the exposition of a natural system of morality, whereas Gould insists that morality is beyond the realm of science.

Morris and Sterelny both miss the opportunity to give us a bottom line on this argument. Gould, Lewontin and their followers believe that we should not take the application of evolutionary theory and genetics to human behaviour seriously, for otherwise we will see a resurgence of eugenics reminiscent of the Holocaust. Their fears may be correct. But no data on brain physiology, no studies on parallel evolution or rapid speciation, and no computer modelling

of complex systems will ever change such perceptions.

These two slim and readable books target a well-defined problem in evolutionary theory. Sterelny could have accomplished more with an index, and both books could have profited from a thorough and organized bibliography. I found it easier to dig up my 20-year-old photocopy of the "Spandrels" article than to find a complete reference to it in either book. Both authors promise an unbiased summary of the arguments, but both come down predominantly in favour of Dawkins' perspective. Morris and Sterelny are on the cusp of an insightful analysis but never quite get to it. But the aficionado of evolutionary theory and the intense debate it engenders would do well to read both accounts. Whereas Morris stresses the divergent approaches of complexity and reductionism, Sterelny emphasizes other issues, such as common descent (or cladistics), which concerns Dawkins, and morphological similarity, which, to Gould, is of paramount importance. Longer, and with an elementary introduction to evolutionary science, Morris's book provides more of a stand-alone account and is suited to the non-specialist.

We have created an icon in Darwin, a god whose every printed word is canon. But Darwin knew that not everything he said would stand the test of time and new data in every detail. Darwin would be puzzled over the struggle for his soul, because the soul, like science, derives its strength not from rigidity but from fluidity. While some of today's most brilliant thinkers grope for the soul of Darwin, it is fortunate that so many experimental evolutionary biologists have decided not to wait for the resolution.

Michael A. Goldman is in the Department of Biology, San Francisco State University, San Francisco, California 94132-1722, USA.

Preaching to the chemical converts

Stories of the Invisible: A Guided Tour of Molecules
by Philip Ball
Oxford University Press: 2001. 195 pp.
£11.99 (pbk)
John Emsley

Science Year has been launched in the United Kingdom. Hallelujah! But will it lead to a revival in interest in chemistry? The irony is that, although the public accepts the tangible benefits

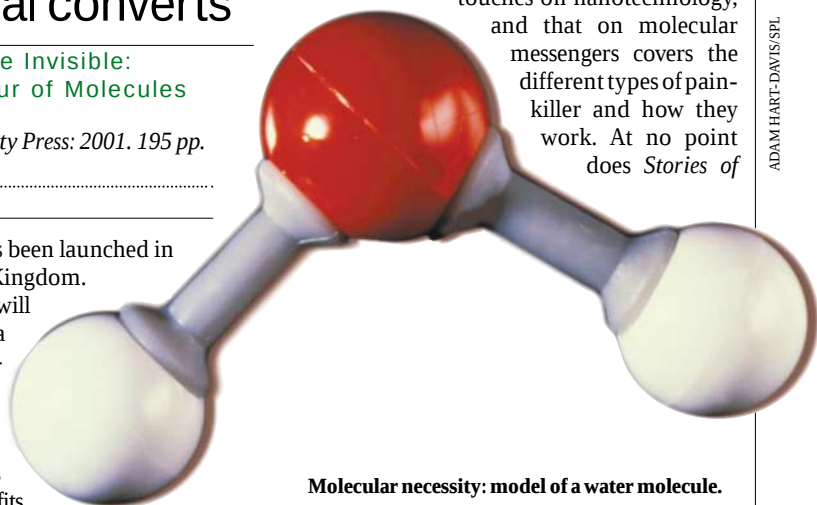
that chemistry delivers, it shows little desire to understand how it works its magic, or to encourage the young to join the faith — a bit like religion, perhaps. In a society of chemical agnostics, it is a brave missionary who tries to reveal its mysteries, but that is what the author of *Stories of the Invisible* has attempted to do — and done remarkably well.

Philip Ball has taken upon himself the task of explaining to the layperson the theories that determine the behaviour of molecules — not just simple molecules, such as water and alcohol, but the complex array of molecules that make up the living cell and those that affect human behaviour. Ball is the right person to write this gospel, and it joins a canon of his successful popular works, the last one of which was the widely acclaimed *H₂O: A Biography of Water*.

Ball knows how to grab people's attention — witness his popular bangs-and-flashes lectures — but a book is different. Can he hold the reader's attention while he explains the intricacies of covalent bonding, stereochemistry, entropy, polypeptides and neurotransmitters? Maybe — the reward for hacking through the thicket of theory is reaching the tree of knowledge, and the attentive reader will manage this feat.

Ball begins with the analogy of letters combining to make words to explain how atoms combine to make molecules. This analogy cannot be stretched too far, but he has found one of the best ways of introducing the concepts of isomerism to a non-chemical audience. The book then moves on to subjects that are bound to capture the reader's attention: what is life? What makes it possible? What keeps a cell alive? How is a cell controlled and how does it store and use information?

These topics are clearly explained, but the book is not solely devoted to explaining the chemistry of living cells and organisms — it also deals with more conventional areas of chemistry. For example, the chapter on energy has a section on explosives, the chapter on organized molecular motion touches on nanotechnology, and that on molecular messengers covers the different types of painkiller and how they work. At no point does *Stories of*



Molecular necessity: model of a water molecule.

ADAM HART-DAVIS/SPL

the Invisible sacrifice sound science for sound bites — we are in the hands of a scholar and a true believer.

Chemistry is undergoing a renaissance as it enters the twenty-first century, and it has become apparent that molecules are the key to understanding other sciences such as medicine, biology, microelectronics and materials. The media will no doubt proclaim wonders in these areas in Science Year, and chemistry will perhaps, indirectly, make some progress in capturing minds, if not hearts.

Ball's heart is clearly in the right place, as his final sentences show: "It is often said that each age tends to interpret the world through models derived from its most advanced technology, and so maybe in the Age of Information we should be wary of becoming too dogmatic about ... questions that haunted Haldane, Schrödinger and countless others. It is perhaps more important that we regard this as a demonstration of the fabulously dynamic, interactive world inhabited, unseen and too often unsung, by molecules." ■

John Emsley is science writer in residence in the Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.

New in paperback is Philip Ball's *The Self-made Tapestry: Pattern Formation in Nature* (Oxford University Press, £12.50, \$19.95).

Reading the history of humanity

The Seven Daughters of Eve: The Astonishing Story That Reveals How Each of Us Can Trace Our Genetic Ancestors/The Seven Daughters of Eve: The Science That Reveals Our Genetic Ancestry by Bryan Sykes
Bantam Press/W. W. Norton: 2001. 320 pp.
£18.99/\$25.95

Howy Jacobs

Mitochondrial DNA represents a unique component of the genome. It is inherited in a strictly maternal fashion, and evolves much faster than nuclear DNA. These properties make it ideal for tracing the history of human populations, especially over the past 100,000 years, during which our species has gained effective mastery of the planet. Untainted by the informational scrambling implicit in sexual recombination, the mitochondrial DNA (mtDNA) represents, as Brian Sykes engagingly explains, a historical record of humanity, a silent witness to our migrations and tribulations.

The best part of the book is the first two-thirds, in which Sykes describes his



Ancient probings: DNA from the fossil bone of Neanderthal hominids may clarify our genetic origins.

endeavours in the field in a style rather like that of a detective story. He eschews strict chronological order (whether of human history or bench science), but presents the research in a logical manner. Human-interest chapters recording the key discoveries are interspersed with well-crafted explanations of basic concepts in cell biology, genetics and evolution. Apart from a few minor blemishes, such as his odd description of mtDNA as "a" gene, *The Seven Daughters of Eve* covers these topics in a way that is both scientifically accurate and understandable to the layperson.

In this first section, Sykes focuses mainly on his own contributions. These include the quest to isolate DNA from ancient specimens such as the celebrated Alpine Iceman, the grizzly work of identifying the remains of the Russian imperial family, the puzzle of where the Pacific islanders came from, the mysterious disappearance of the Neanderthals and, finally — the main story of the book — the ancestry of modern Europeans.

Sykes acknowledges his debt to others in the field — pioneers such as Rebecca Cann and Svante Pääbo, and co-workers such as Chris Stringer. On the other hand, Luca Cavalli-Sforza may find the chapter entitled "We are not amused", which deals with the long-running academic feud over European origins, rather more vexatious. However, Cavalli-Sforza finally agreed with Sykes that the gene pool of present-day Europeans comes predominantly from Palaeolithic hunter-gatherers, with only a minor contribution from early farmers who migrated into Europe during the Neolithic period. Sykes is gracious (in victory), however, and many other scientists will relate to the emotionally vivid way he portrays this battle of ideas.

The last part of the book, from which the title derives, is, correspondingly, slightly

disappointing. I found the accounts of the imagined lives of the seven founders of the major European mtDNA lineages (haplogroups) rather banal, at times verging on the puerile. The tales are marred by unfortunate anachronisms, such as the description of a successfully crafted flint tool as "a real peach", and simple inaccuracies, such as the ancient Mediterranean tide coming in. Sykes laces the stories with somewhat forced descriptions of the invention of agriculture, the development of ancient trading practices, cave art and the first boats. All this made me feel that I was reading someone's school project, with influences from *The Flintstones* cartoon series, rather than a treatise by a leading academic. I would have preferred a shorter account, based on specific archaeological findings, in the same style as the earlier chapters. But some readers may enjoy this semi-fictional diversion.

Sykes pours deserved scorn on the primitive genetic theories that inspired nineteenth- and twentieth-century racists. Yet the romanticism of his accounts of the seven clan-mothers comes rather close to achieving the opposite effect. I began to wonder whether this was leading up to a proposed resolution to the problem of the inexorably growing European Commission: instead of one commissioner from each member state, why not one from each of the seven matrilineal clans? Maybe that's not such a bad idea.

Despite these minor faults, and despite more academic criticisms that one might make of his data interpretation, this will be recognized as an important work, bringing molecular anthropology to a mass audience. ■

Howy Jacobs is at the Institute of Medical Technology, FIN-33014 University of Tampere, Finland.

Sex, love and science

Serious treatments of science in fiction can be illuminating for both scientists and non-scientists.

Susan Gaines

A nineteenth-century woman naturalist burns with thwarted scientific curiosity, revelling in a clandestine experiment. Three physicists working to reconcile quantum mechanics and relativity form a tragic triangle of erotic, filial and intellectual love. Another solves the problem he most wanted to solve, but the answer, the fame and his solicitous lover all fail to satisfy his longings. Astronomy expands and nurtures the inner life of a man so ugly that he worships beauty as if it were a transcendent god.

These tales of sex, love and science come from fiction — a story from Andrea Barrett's *Ship Fever*, Lisa Grunwald's novel *The Theory of Everything*, Rebecca Goldstein's *Properties of Light* and Chet Raymo's *Dork of Cork*. In recent years, novelists have been mining science for metaphors, faddishly sprinkling literature with scientific allusions, but these books go beyond allusion to focus directly on the scientific enterprise. Not science fiction, with its emphasis on futuristic technologies, or *Jurassic Park*-style commercial entertainment, these are literary explorations of scientific culture, featuring characters whose lives are informed by direct involvement with scientific knowledge.

Forty years after C. P. Snow sounded an alarm about the cultural divide between science and the humanities, with Alan Sokal only recently taking post-modern philosophers to task for their abuse of physics, a few novelists — some of them also scientists or former scientists — have been approaching science with their literary inquiries. Whether Snow's two cultures have fused, or fragmented, or been obliterated by commercialism and technology, the rift between scientists and non-scientists remains — not for the inevitable gap in their knowledge, but for the way they perceive knowledge —

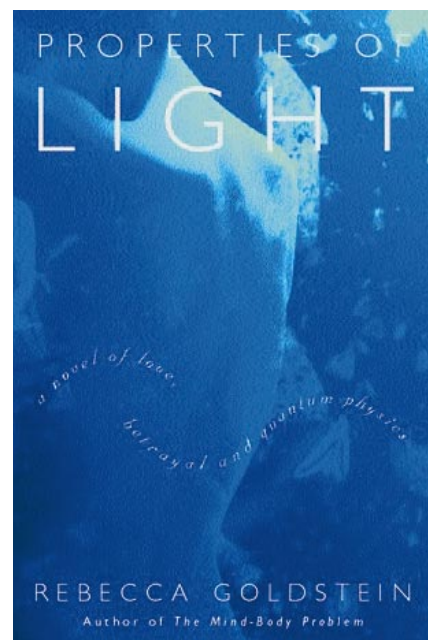
“See,” these novels seem to tell us, “science is common knowledge; science is the stuff of life and literature.”

and novels such as these may play a small but interesting role in closing it. They give a daringly intimate, idiosyncratic, irreverent, passionate and honest view of science and its personal and philosophical dilemmas, a perspective that is rarely seen in public discourse.

In Anna McGrail's *Mrs Einstein*, the illegitimate daughter of a famous physicist becomes obsessed with her father's refusal to acknowledge her and twists her mathematical talent into a vehicle for revenge. In Louis B. Jones's *Particles and Luck*, a physicist bumbles through a suburban comedy of errors, and yearns for “the ordinary manly pleasure of certainty” and “a job to do” rather than a life spent in “scrupulous wimpy professional self-doubt”. With the First World War brewing, two naturalists struggle to resolve the paradox of human altruism and natural selection in Michel Rio's novella *Dreaming Jungles*. Richard Powers' *The Gold Bug Variations* spins a metaphorical web of love, music and 1950s molecular biology, while affirming the infectious power of scientific curiosity. Rick Bass's *Where the Sea Never Ends* tells of the relationship between a petroleum geologist, the land he's exploring, and the ageing oil magnate who mentors him.

The language and culture of science, which are designed to nullify the personal and subjective, do not accommodate such themes easily. A vocabulary of unequivocal definitions, proclivity for the passive tense, and avoidance of the first-person singular serves the objective treatment of science well. But when it comes to intersections between science and values, both public and personal, the unbridled poetry and literary sensibility of a novelist can be as welcome as a mountain breeze wafting through a hermetically sealed building. Recently, after a presentation of my novel about a geochemist, a carbon-cycle researcher in the audience said to me, “It's hard to express this stuff”, and added, “Actually, most of the scientists I know are knee-jerk environmentalists”. He was talking about their personal reactions to their data as a thing apart from the data or their production, referring to the scene I'd read. It was treacherous, oft-avoided territory for scientists, but as a novelist I'd marched right in, and he seemed grateful for the excuse to explore it.

And what of the data in these novels? What about the science itself? Can we have meaningful discussions of scientific knowledge in fiction? Simon Mawer meets this



Matter of the heart: love, intellectual envy and fundamental questions of physics are combined in Rebecca Goldstein's *Properties of Light*.

challenge head-on in *Mendel's Dwarf*, the story of a geneticist whose scientific outlook is forged by the overriding fact of his life — his physical deformity — and whose view of the world is so dominated by his knowledge of genetics that his thoughts come replete with journal references. Mawer's novel does what many popular science books and textbooks fail to do in their rush to serve up the facts — it dips beneath the smooth, unrippled surface of established knowledge to the wandering lines of thought and experiment that support it, the rationale for the facts.

Chemist Carl Djerassi has coined the term ‘science-in-fiction’ for his own novels, and states clearly that his goal is to educate. The literary novel is, by its nature, non-didactic; not a lesson in science, nor one in morality, it can only be an elucidation of either, or of both. Ironically, it is precisely this quality that can serve Djerassi's purpose, in ways that other forms of discourse do not. The novels discussed here are not attempts to disseminate scientific knowledge, to popularize science or to sensitize scientists. They belong to no genre or school, vary wildly in style and theme, and make their own rules about the depth and rigour with which they discuss science. But implicit in this small body of literature is the assumption — the presumption — that science is approachable. “See,” these novels seem to tell us, “science is common knowledge; science is the stuff of life and literature.” And perhaps, by presuming so, they nudge us closer to making it true. ■

Susan Gaines crossed over from science to fiction more than a decade ago. She is the author of the novel *Carbon Dreams* (Creative Arts Book Co.).

A principled stance

Martyn Poliakoff and Paul Anastas

In Chinese, chemistry is the 'mixing science', whereas in Dutch, it's the 'separation art'. But whatever their nationality, academic chemists often do not know in detail how the chemicals they use are made, or how their chemistry affects the biosphere. They view industrial chemistry as meat-eaters view the slaughterhouse — the gory details are glossed over and the fate of the waste is ignored. On the other hand, much of the general public recoils at the mere mention of 'chemicals', yet we are all clothed, fed, washed and transported by the products of the chemical and pharmaceutical industries. While we benefit from increased quality and length of life thanks to chemicals, the industry that makes them is frequently accused of degrading the Earth.

Society relies on chemicals. Like a good cook, the industrial chemist recycles waste. Yet many types of waste are dangerous, leaving no option but costly disposal. With the global population rising and standards of living increasing, current methods of chemical production are unsustainable. As production rises to meet demand, waste levels will soar and landfill sites will be exhausted. Manufacturers will be increasingly restricted by environmental legislation; enforcement agencies will become increasingly overloaded; and costs of waste treatment will stifle innovation.

A new approach is needed to remould attitudes constructively, and to attract young people to the field of chemistry. Step forward 'green chemistry', which aims to create products that are as harmless as possible and therefore require less regulation. The concept is enshrined in a set of 12 principles, ranging from minimizing waste to avoiding accidents. Individually, these principles are not new; what is potentially revolutionary is grouping them together so that chemists can focus on using classical chemistry to design products with less environmental impact. For example, non-toxic carbon dioxide is now used to make lighter, stronger building materials and to replace chlorinated organic solvents in chemical reactions and dry-cleaning. Integrating energy, waste and toxicity issues when thinking about new processes and reactions has proved to be very effective.

Although some industrial processes are efficient, others are extremely wasteful, requiring costly handling and disposal of chemicals. The more stages a chemical process involves, the more potential it has for creating waste. So green chemistry inspires a kind of chemical 'golf match', in which fewer steps represent a better environmental 'score' — and the best result is, of course, a hole in one. For example, a new process for producing the anti-inflammatory drug ibuprofen halves the number of steps, makes them all catalytic and more than doubles the atom efficiency of the process. Green chemistry not only leads to cleaner and more efficient processes, but can increase profitability by eliminating many of the traditional costs of treatment, disposal, liability and regulatory compliance.

Green chemistry involves more than just tidying up existing processes — if it were that simple, process economics would already be driving manufacturers in that direction. Rather, it identifies a need to design new, safer molecules. Often, the toxicity and the usefulness of a compound arise from different parts of the molecule; for example, a couple of strategically placed methyl groups added to a dye molecule makes it much less toxic but leaves its dyeing properties intact. Similar minor structural modifications can increase the biodegradability of molecules and reduce their environmental impact. Linking structure and toxicity gives synthetic chemists a new perspective, allowing them to design new pesticides that are toxic only to target organisms, with a fraction of the usual dose, and which do not persist in the environment.

Green chemistry has captured the imagination of many chemists. But some industrial chemists believe that their very successful efforts at cleaning up have been overlooked, and some academics fear that green chemistry

Green chemistry

Already, young people are beginning to perceive chemistry as a key to saving our environment rather than a tool for its destruction.

is 'soft' science. Although understandable in historical contexts, neither view is correct. Green chemistry is a new partnership, bringing together the efforts of industrial and academic research and building rapidly on the past successes of both. New solvents, feedstocks, catalysts and processes, producing everything from new, biologically derived, renewable plastics to lead-free automotive paint, are being developed throughout the world.

Many challenges remain. How can chemical production shift from unsustainable petroleum feedstocks to renewable biomass? Can some of the vast quantities of carbon dioxide vented into the atmosphere be converted into useful chemicals? Can the environmental impact of agrochemicals be reduced while increasing food production? Green chemistry will attract the new generation of chemists needed to solve these problems. Already, young people are beginning to see chemistry as a key to saving our environment rather than a tool for its destruction.

Will green chemistry take over all of chemistry? Far from it — the revolution of one generation becomes the orthodoxy of the next. The 12 principles are so obvious that chemists of the future will wonder why it took so long to integrate them into the mainstream. After all, why make chemicals wastefully and expensively, when they can be produced cheaply and cleanly? ■

Martyn Poliakoff is a research professor and Paul Anastas is a special professor at the School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, UK.

FURTHER READING

- Anastas, P. & Warner, J. *Green Chemistry: Theory and Practice* (Oxford Univ. Press, London, 1998).
Anastas, P. & Lankey, R. *Chem. Britain* **37**, 26–29 (2001).
Hamley, P. A. & Poliakoff, M. *Chem. Engineer* **72**, 24–27 (2001).
Ritter, S. K. *Chem. Eng. News* **79**, 24–37 (2001).

WEBLINKS

- Green chemistry network:
♦ <http://www.rsc.org/networks/gcn>
What is green chemistry?:
♦ <http://www.epa.gov/opptintr/greenchemistry/whatis.htm>
Green chemistry at Nottingham:
♦ <http://www.nottingham.ac.uk/greenchemistry>



Green future: the world needs cleaner chemistry.

Walking with whales

Christian de Muizon

Whales must have evolved from land-based mammals, but fossil evidence of some of the steps in between has been patchy. Newly discovered skeletons with legs fill in the gaps.

On page 277 of this issue, Thewissen and colleagues¹ describe their discovery of partial fossil skeletons from the earliest cetaceans — a group of mammals that today consists of whales, porpoises and dolphins. The fossils, which are some 50 million years old and were found in Pakistan, take us a huge step forwards in understanding the origins and evolutionary relationships of whales. Until now, the limbs of all known early cetaceans reflected an amphibious or wholly aquatic lifestyle^{2–6}. But the newly discovered fossils show that the first whales were fully terrestrial, and were even efficient runners. They also reveal that cetaceans are more closely related to the oldest known even-toed ungulates — a group of hoofed mammals that includes cows, hippos, pigs, camels and giraffes — than to any other mammals. These conclusions are based on solid anatomical data, and contradict the previous hypotheses of both palaeontologists and molecular biologists^{7–10}.

All of the mammals that existed in the early Tertiary — some 65–50 million years ago — lived on land. So it has always been clear that aquatic cetaceans must have evolved from terrestrial mammals and returned to the water, and the forelimbs of recent cetaceans still have the same general pattern as that of land mammals. However, modern cetaceans are highly specialized, with numerous adaptations that allow them to swim, dive and feed. Their bodies are elongated and streamlined; the shape of the tail is modified into a propulsive fluke; and the limbs are much reduced in size. The forelimbs are flippers with rigid elbows and an increased number of phalanges (part of the finger bone), and function mainly in steering and stabilizing. The hindlimbs and pelvis are internal, and have no role in swimming.

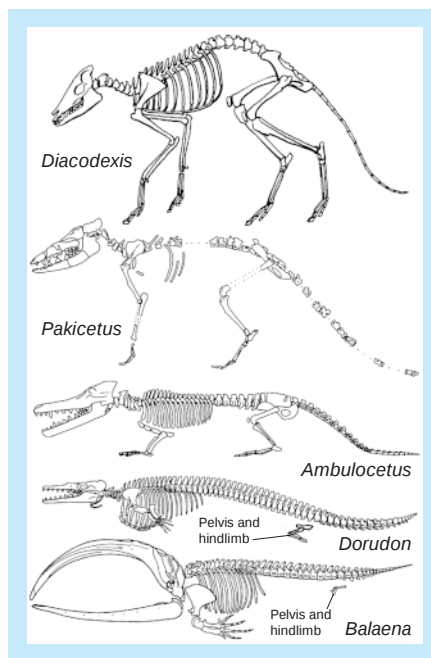


Figure 1 The evolutionary route to a whale. *Diacodexis*¹³ was a primitive even-toed ungulate (hoofed mammal); *Pakicetus* is one of the terrestrial cetaceans described by Thewissen *et al.*¹; *Ambulocetus*¹⁴ was amphibious; *Dorudon* (modified from ref. 15) was a fully aquatic archaeocete (early cetacean), but retained an articulated elbow and vestigial hindlimbs; and *Balaena* is a recent whale¹⁶. Skeletons are not drawn to scale.

What was the evolutionary path from the earliest land mammals to modern, aquatic whales? Pakicetids, which existed in the early Eocene epoch, some 50 million years ago, are the earliest known cetaceans. But, until now, only the skulls of these animals had

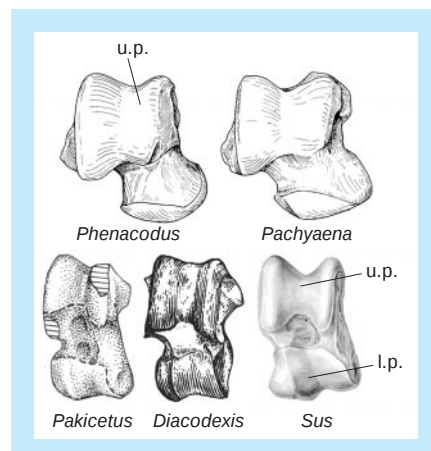


Figure 2 Ankle bones (astragali) of hoofed mammals and primitive whales. *Phenacodus*¹⁷, a primitive ungulate, had an unspecialized ankle bone that resembles that of *Pachyaena*¹⁷, a mesonychian ungulate from 50 million years ago. The double-pulleyed astragali of *Pakicetus* (one of the fossil cetaceans described by Thewissen *et al.*¹), *Diacodexis*¹⁸ (the oldest known even-toed ungulate), and *Sus*¹⁷ (the pig) indicate a close relationship between these species. Bones are not shown to scale. u.p., upper pulley, articulates with tibia. l.p., lower pulley, articulates with distal ankle bones.

been described. Cetaceans from the middle Eocene, 45 million years ago, have been known for more than a century¹¹, but here, too, few limb remains were discovered, until about 20 years ago. Since then, numerous fossils from North America, Pakistan and Egypt have revealed that these early cetaceans had mobile elbows and external hindlimbs with articulated knees^{2–7} (Fig. 1). However, they were already fully aquatic, except for

Box 1 Further finds from Pakistan

A report in this week's *Science* by Gingerich *et al.*¹⁹ fleshes out the picture of early whale evolution with a description of two new species from 47-million-year-old rocks in Pakistan. One represents a new genus; the other, a close relative, is a species of the

previously known genus *Rodhocetus*. Both belonged to the family Protocetidae, a group of extinct whales somewhat more aquatic in their adaptations than the pakicetids described elsewhere in this issue by Thewissen and colleagues¹. Like the previously

described *Ambulocetus*, the newly discovered protocetids had well-developed limbs with large hands and feet. Gingerich *et al.* suggest that although these creatures could have moved about on land, the limbs would have functioned as paddles. In

life, the animals, which would have weighed between 400 and 500 kilograms, might have lived and moved like large sea lions. The most striking features of these protocetids are their astragali, which look remarkably like those of

artiodactyls and reaffirm the close relationship between artiodactyls and whales. However, the absence of a formal phylogenetic analysis means that the researchers were unable to test the competing hypotheses of whale relationships. **Henry Gee**

Ambulocetus, which was amphibious — much like sea lions. So there was no detailed information about the anatomy of the cetaceans' terrestrial ancestor.

Recent cetaceans are very different to other mammals, so another question that has dogged this field is that of which group of mammals contains their closest relatives — which is their 'sister group'? The cranial and skeletal anatomy of cetaceans is highly modified compared with that of land mammals, and fossils of early cetaceans are so rare and generally incomplete, that the affinities of the group are difficult to establish. On the basis of tooth and ear morphology, palaeontologists contend that cetaceans are most closely related to the mesonychians^{7,8} — a group of extinct ungulates from the early Tertiary. But molecular biologists favour hippos — which form one of the families of modern even-toed ungulates (artiodactyls)^{9,10} — as the sister group.

Thewissen and colleagues' discovery¹ allows us to address both of these problems. The newly found fossils include several skulls and postcranial bones from two early pakicetid species — which, it seems, had the head of a primitive cetacean (as indicated by the ear region) and the body of an artiodactyl. All the postcranial bones indicate that pakicetids were land mammals, and it is likely that they would have been thought of as some primitive terrestrial artiodactyl if they had been found without their skulls. Many of the fossils' features — including the length of the cervical vertebrae, the relatively rigid articulations of the lumbar vertebrae, and the long, slender limb bones — indicate that the animals were runners, moving with only their digits touching the ground.

But the most eloquent information provided by the fossils¹ comes from the ankle bones, particularly the astragalus (Fig. 2). This has two pulleys, which connect to both the tibia and the more distal ankle bones and allow a great deal of flexibility. This type of morphology is an adaptation for running. It was once thought to be unique to artiodactyls, but it is now clear that it also occurred in cetaceans. So, for the first time, morphological evidence shows that artiodactyls are the closest relatives of the cetaceans (also see Box 1).

This means that artiodactyls and cetaceans form the two branches of a larger group, the cetartiodactyls¹². Thewissen and colleagues' studies also exclude the mesonychians from this larger group, in part because these ungulates do not have a double-pulleyed astragalus. Mesonychians are an unspecialized group of primitive ungulates, and perhaps some of them should be included in other groups of mammals. Cetaceans and some mesonychians have dental similarities and an elongated skull, but these features are probably the result of convergent evolution.

The authors' analysis also indicates that

no one recent artiodactyl family is more closely related than another to cetaceans — in other words, hippos are not the extant sister group of the cetaceans. The closest fossil relatives of the cetaceans were probably the earliest known artiodactyls, such as *Diacodexis* (see Fig. 1).

Thewissen *et al.*'s discovery¹ of these terrestrial cetaceans is one of the most important events in the past century of vertebrate palaeontology. Only a very few fossils, such as these, reveal a link between two groups of vertebrates that are hugely different in terms of shape yet closely related in terms of evolution. When there is a drastic shift in habitat — such as from land to water — the morphology of the newly adapted animals is generally so greatly modified, because of the high selective pressure, that any resemblance to the original ancestor is quickly obliterated. But the new fossils¹ superbly document the link between modern whales and their land-based forebears, and should take their place among other famous 'intermediates', such as the most primitive bird, *Archaeopteryx*, and the early hominid *Australopithecus*. ■

Christian de Muizon is at the Unité Mixte de Recherche 8569, Centre National de la Recherche Scientifique, Laboratoire de Paléontologie, Muséum

National d'Histoire Naturelle, Paris 75005, France. e-mail: muizon@cimrsl.mnhn.fr

1. Thewissen, J. G. M., Williams, E. M., Roe, L. J. & Hussain, S. T. *Nature* **413**, 277–281 (2001).
2. Gingerich, P. D., Smith, B. H. & Simons, E. L. *Science* **249**, 154–157 (1990).
3. Gingerich, P. D. *et al.* *Nature* **368**, 844–847 (1994).
4. Thewissen, J. G. M. *et al.* *Science* **263**, 210–212 (1994).
5. Thewissen, J. G. M., Madar, S. I. & Hussain, S. T. *Cour. Forschungsinst. Senckenberg* **191**, 1–86 (1996).
6. Uhen, M. D. in *The Emergence of Whales: Evolutionary Patterns in the Origin of Cetacea* (ed. Thewissen, J. G. M.) 29–61 (Plenum, New York, 1998).
7. Geisler, J. H. & Luo, Z. in *The Emergence of Whales: Evolutionary Patterns in the Origin of Cetacea* (ed. Thewissen, J. G. M.) 163–212 (Plenum, New York, 1998).
8. O'Leary, M. A. & Geisler, J. H. *Syst. Biol.* **48**, 455–490 (1999).
9. Shimamura, M. *et al.* *Nature* **388**, 666–670 (1997).
10. Nikaido, M., Rooney, A. P. & Okada, N. *Proc. Natl Acad. Sci. USA* **96**, 10261–10266 (1999).
11. Kellogg, R. A. *Review of the Archaeoceti* (Carnegie Inst., Washington, DC, 1936).
12. Montgelard, C., Catzeflis, F. M. & Douzery, E. *Mol. Biol. Evol.* **14**, 550–559 (1997).
13. Janis, C. M., Scott, K. M. & Jacobs, L. L. (eds) in *Evolution of Tertiary Mammals of North America* p. 360 (Cambridge Univ. Press, 1998).
14. <http://www.neoucom.edu/Depts/Anat/Ambulocet.html>
15. Gingerich, P. D. & Uhen, M. D. *Contrib. Mus. Palaeontol. Univ. Michigan* **29**, 359–401 (1966).
16. Eschricht, D. F. & Reinhardt, J. in *Recent Memoirs on the Cetacea* (ed. Flower, W. H.) 1–150 (Ray Soc., London, 1866).
17. O'Leary, M. A. & Rose, K. D. *J. Vert. Paleontol.* **15**, 401–430 (1995).
18. Schaeffer, B. *Am. Mus. Nov.* **1356**, 1–24 (1947).
19. Gingerich, P. D., ul Haq, M., Zalmout, I., Khan, I. H. & Malkani, M. S. *Science* **293**, 2239–2242 (2001).

Surface physics

A new crack at friction

David A. Kessler

One of the dirty little secrets of physics is that there is no generally accepted explanation of the basic laws of friction. An advance in the theory of cracks will stimulate fresh thinking on the question.

Friction is a ubiquitous feature of everyday life. Without it, we couldn't walk, tyres wouldn't roll, and ballpoint pens would fail to write. But what is friction, and how does it act?

The basic properties are simple to grasp. To move a solid object from rest on top of a solid surface, a minimum force has to be applied to overcome the force of friction. This force is proportional to the compressive force pushing the two surfaces together, in this case the weight of the object. Intriguingly, this minimum force is independent of the area of contact between the body and the surface. So the friction force on a rectangular solid resting on a table is the same whichever face is in contact with the surface. These laws have been known since the mid 1700s and are attributed to the French physicists Guillaume Amontons (1663–1705) and Charles Augustin de Coulomb (1736–1806). It is one of the dirty little secrets of physics that while we physicists can tell you a lot about quarks, quasars and other exotica, there is still no universally accepted explanation of the basic laws of friction. On page

285 of this issue, Gerde and Marder¹ offer a theory of surface cracks that may lead to a better understanding of surface friction.

The standard picture of friction² (dating from as recently as the 1960s) is that the solid surfaces are not really planar, but are rough on a microscopic scale. The presence of these tiny surface features, or asperities as they are known, prevents the surfaces from coming into full contact. So the true contact area is much smaller than its apparent value, and is proportional to the compressive force between the surfaces, in much the same way that the contact area between a car tyre and the road increases when you load your car. Problems have arisen when physicists tried to confirm this picture using calculation from first principles. The goal is to construct, either analytically or on the computer, a solid body and surface from atoms with prescribed interactions, and calculate the friction force directly. But previous attempts at this found that the two surfaces ride freely on top of each other because of the mismatch between the asperities on the two surfaces, so there is no friction.

Two alternative solutions to the problem have recently been proposed. The first, suggested by Müser, Wenning and Robbins³, attributes a crucial role to dirt — the diffuse collection of foreign mobile atoms trapped between the two surfaces. According to the authors' numerical simulations, these mobile atoms quickly find appropriate gaps between the surfaces where they become trapped. These atoms then 'lock' the two surfaces in place. To move the top surface, it has to be pushed up and over the dirt atoms, the force required being proportional to the weight of the top body. Furthermore, the calculated force is seen to be essentially independent of the apparent contact area.

Gerde and Marder¹ suggest a second, radically different mechanism, based on the physics of self-healing shear fracture. The basic idea is simple. Imagine you want to move a large rug some distance along the floor. Instead of dragging it, a less back-breaking method is to lift the back edge, slide the edge forward a bit and so introduce a ridge in the rug. Pushing on the ridge moves it forwards along the length of the rug until it reaches the end. The net result is that the rug has been shifted. If one pushed uniformly on the rug and not just on the ridge (a uniform shear load), the ridge would move on its own. In essence, this is a kind of surface fracture; the two surfaces at the ridge (the rug and the floor) are now no longer in contact, just as the two halves of a material break contact and come apart at a crack.

This sort of surface fracture is different to the type created when two surfaces are pulled apart (called a mode I crack). In this case, the surfaces are kept far apart, whereas in the fracture considered by Gerde and Marder (a mode II crack) the two surfaces remain in close proximity and can come together again and bond, or 're-heal', in their new laterally shifted positions. This rebonding sets in some distance downstream of the crack edge, so the crack is of finite length. Such self-healing cracks have been suggested to play a crucial role in the dynamics of earthquakes, and apply wherever two surfaces are sheared relative to one another. The existence of such cracks could help to explain the low amount of heat generated by some earthquakes.

The classic theory of self-healing cracks has been plagued by problems. In the standard mathematical treatment, the two surfaces are predicted to oscillate an infinite number of times, with decreasing wavelength as one approaches the edge of the crack. The meaning of this mathematical nightmare has been unclear, and is the subject of considerable controversy. The situation is reminiscent, though in some sense worse, than the situation for mode I cracks.

In the mode I case, where the surfaces are pulled apart, the classic theory predicts infinite stress at the edge of the crack. This problem was resolved by the work of Leonid

Slepyan^{4,5}, who used an atomic lattice model to study fractures. Using Slepyan's analytical solution of the problem, it can be shown⁶ that in the atomic lattice model the stresses increase in accord with the standard theory as the crack edge is approached, until just a few lattice spacings from the edge. At this point, the true stress saturates, deviating from the predictions of the standard theory. Furthermore, the atomic-lattice calculation predicts the velocity of a propagating crack as a function of the applied stress — a factor that has to be taken from experiments in the standard treatment. Slepyan's theory accomplishes for mode I cracks exactly what we want from a theory of friction: it tells us how to get from the microscopic force laws to a macroscopically measurable quantity, here the crack velocity.

In a mathematical tour de force, Gerde and Marder¹ have succeeded in extending the Slepyan model to a mode II crack of finite length. Here, too, the atomic-lattice solution succeeds where the standard theory fails (at the crack edge). Their model also predicts the minimum shear force required to initiate a propagating crack. Gerde and Marder find that this minimum shear force is roughly

proportional to the compressive force pushing the surfaces together. In other words, for the surfaces to move relative to one another, the applied force has to be proportional to the compressive force, as in the Amontons–Coulomb law. The idea then is that, once this minimum force is applied, self-healing cracks are created, causing the surfaces to slide past one another. Gerde and Marder's work is a significant step forward in the theory of cracks. Whether it also helps to solve the problem of friction, or whether some other theory along the lines of that put forward by Müser, Wenning and Robbins³ will prove more fruitful, remains to be seen. It can be hoped that these studies will stimulate new experiments to decide the issue. ■

David A. Kessler is in the Department of Physics, Bar-Ilan University, Ramat-Gan, Israel.

e-mail: kessler@dave.ph.biu.ac.il

1. Gerde, E. & Marder, M. *Nature* **413**, 285–288 (2001).
2. Bowden, F. P. & Tabor, D. *The Friction and Lubrication of Solids* (Clarendon, Oxford, 1986).
3. Müser, M. H., Wenning, L. & Robbins, M. O. *Phys. Rev. Lett.* **86**, 1295–1298 (2001).
4. Slepyan, L. I. *Izv. Akad. Nauk SSSR Mekh. Tverd. Tela* **16**, 101–115 (1982).
5. Kulamekhtova, Sh. A., Saraikin, V. A. & Slepyan, L. I. *Izv. Akad. Nauk SSSR Mekh. Tverd. Tela* **19**, 112–118 (1984).
6. Kessler, D. A. & Levine, H. *Phys. Rev. E* **59**, 5154–5164 (1999).

Plant biology

Mobile protein signals cell fate

Sarah Hake

In plant roots, different cell types are organized in a well-defined pattern: each cell knows exactly where it is and what it should do. A molecule that tells cells where they are has now been discovered.

In plants, the fate of each cell is determined mainly by its location. Plant cells 'know' if they are on the top or bottom of a leaf, the inside or outside of a stem, or near or far from vascular tissue, and they become specialized (differentiate) accordingly. Yet it has not been known which molecules function as positional cues or how these signals are transmitted.

Roots are a handy system for studying cell fate because they have a limited number of different cell types, which form well-defined concentric rings. On page 307 of this issue, Nakajima and colleagues¹ show that a protein known as SHORT-ROOT (SHR) is likely to be a positional cue in the roots of thale cress, *Arabidopsis thaliana*. SHR relays its information by moving from an inner cell layer to cells in the adjacent outer layer. Once there, it enters the nucleus and promotes the expression of a related gene, *SCARECROW* (*SCR*), whose protein product is needed for asymmetric cell division and the differentiation of two particular cell types. The movement between plant cells of proteins that control gene expression (transcription factors) has been seen before,

but this is the first time that a purpose has been identified.

Cell division and differentiation occur in an orderly fashion in the *Arabidopsis* root,

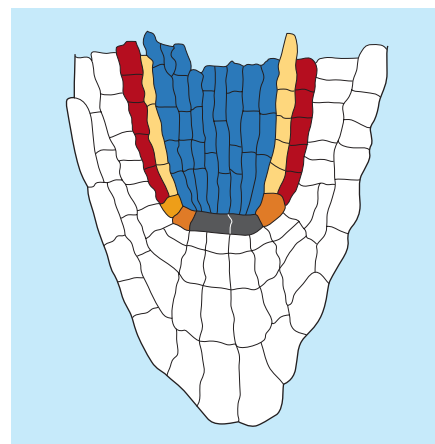


Figure 1 The organization of cell types in the tip of an *Arabidopsis* root. Black, quiescent-centre cells; dark orange, cortex–endodermal initial cells; pale orange, daughter cell of a cortex–endodermal initial; red, cortex; yellow, endodermis; blue, stele. Other cell types are white.

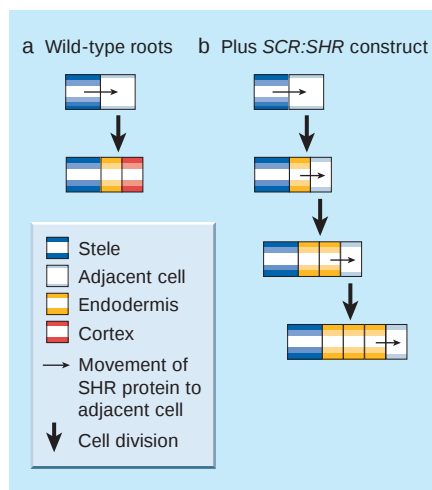


Figure 2 Patterning the roots. a, The differentiation of cells into endodermis and cortex in the *Arabidopsis* root depends on the movement of the SHORT-ROOT (SHR) protein from the stele to the adjacent cell layers. b, Nakajima *et al.*¹ placed the *SHR* gene under the control of the *SCR* promoter region, ensuring that *SHR* is expressed in the layers adjacent to the stele. The result is the formation of many layers of endodermis.

making it possible for researchers to produce a precise map of division patterns and cell fates (Fig. 1). The lineage of each cell type can be traced back to the 'quiescent centre', found in the growing root tip.

Cells in the quiescent centre rarely divide. But when they do, they produce different types of 'initial cell', which in turn supply more specialized cell types. Quiescent-centre cells divide downwards to produce the root cap, upwards to produce the stele (primary vascular tissues), and sideways to produce initial cells for the other three layers (from the outside inwards: epidermis, cortex and endodermis). The cortex and endodermis are derived from the same initial cell; this cell divides once to make a daughter cell that itself divides asymmetrically to produce a row of cortex and a row of endodermal cells (Fig. 1). One might think that the fate of each cell is determined by these strict patterns of division. However, studies in which different cell types were ablated from roots show that a cell's position is more important than its lineage².

Two genes are required for the differentiation of endodermis and cortex: *SCR* and *SHR*. Plants with mutated *SCR* lack the crucial asymmetric cell division that leads to separate cortex and endodermis. The resulting single-celled layer differentiates into a combination of cortex and endodermis, as judged by the expression of cell-specific markers. In plants with mutations in *SHR*, this cell division again fails and the resulting cell layer lacks endodermal markers but has attributes of the cortex^{3,4}.

In normal plants, the daughter cell pro-

duced by the cortex–endodermal initial expresses *SCR* before it divides asymmetrically. Afterwards, *SCR* is expressed in the endodermal layer^{5,6}. Expression of *SCR* requires *SHR*. But the patterns of *SHR* expression are an enigma — *SHR* is transcribed in the stele, rather than in the endodermis where it presumably functions⁷. This led to speculation that *SHR* functions in a non-cell-autonomous way, meaning that it influences adjacent cells in which it is not transcribed⁷.

Non-autonomous action has been reported for several plant genes⁸. These non-autonomous effects were discovered in experiments involving chimaeric plants, in which adjacent cell layers had different genotypes (the gene of interest might be normal in one layer and mutated in the next). Often, the products of the genes might be produced in only one cell layer and yet be able to 'rescue' adjacent mutant cells. We already know that transcription factors can move between plant cells⁹ through the cell walls, which have small, dynamic pores called plasmodesmata that can open under certain conditions. But the question remained as to why protein movement is necessary, given that the proteins could be — and often are — expressed in both cells in the first place.

Using an antibody that recognizes *SHR*, as well as a fusion of *SHR* and green fluorescent protein, Nakajima *et al.*¹ now show that *SHR* protein is found in the stele, where it is transcribed⁷. It also moves to the adjacent layer, the endodermis. *SHR* is located in the nucleus of endodermal cells, and is found both inside and outside the nucleus in the stele. The authors also detected it in quiescent-centre cells and the cortex–endodermal initial and daughter cells (see Fig. 1b on page 308). These results explain *SHR*'s non-autonomous effects, but do not explain the need for such a system.

So Nakajima *et al.* went on to investigate what happens if *SHR* is expressed in the endodermis, rather than moving there. Using a regulatory region (the promoter) from the *SCR* gene, they drove expression of the *SHR* gene in the endodermis. This resulted in several layers that expressed endodermal markers (Fig. 2). Apparently, if the *SHR* protein is expressed in the endodermis, it moves one layer outwards and then initiates the transcription of *SCR*, which is then required for the asymmetric cell division leading to endodermis and cortex. Once the newly divided cells are specified as endodermis, the expression of *SHR*, under the control of the *SCR* promoter, begins again; the protein moves outwards once more, and so the autocatalytic cycle is perpetuated. The extra layers did not form in plants with mutations in *SCR*. So, proteins that are programmed to move one layer and regulate downstream genes are perfect candidates for positional signals. Directionality



100 YEARS AGO

Dr. Carl Lumholtz, the Norwegian explorer, who for the past five years has been travelling in the hitherto unknown regions of North-Western Mexico for the American Museum of Natural History, lectured before the Geographical Society in Christiania on September 12 and gave a description of his life and travels among the wild Indian tribes of the Western Sierra Madre... In order to study these interesting people he sent back the entire staff of his expedition and lived alone among them. At first the tribes objected to his taking up his abode in this way, but eventually he gained their confidence and was allowed to remain. He learnt their ways, their language and their songs, and joined in their dances... They are very intellectual, and are, according to Dr. Lumholtz, a far superior race to their kinsmen in the United States and South America. Among many of the tribes he found a higher degree of morality than in civilised countries.

From *Nature* 19 September 1901.

50 YEARS AGO

It is impossible to predict precisely the course of a great aurora, but what usually happens is as follows. At first, a glow, much like dawn (hence the name aurora borealis, northern dawn, which was first used by the French philosopher, Gassendi, in 1621), appears on the northern horizon... The glow then ascends from the horizon to form an arc, the lower border of which is at a height of about 100–110 km... The true colour of most arcs is probably green, for a predominant line in the spectrum is the green forbidden line 5577 Å. of atomic oxygen, and this happens to be in the region of maximum sensitivity of the retina. But, except in the case of very bright arcs, it appears to the eye to be grey-white, since the light intensity is below the threshold of colour perception. The arc may then brighten suddenly to an intensity above the threshold such that patches of vivid greens and reds can be seen here and there along its length. Simultaneously the lower border becomes extraordinarily sharply defined and is appreciably lower, all of which are signs that the arc is about to lose its regular shape and to show ray structure. It may then undulate to appear like the folds of a huge waving curtain, or it may take the form of a wide band, stretching irregularly across the sky.

From *Nature* 22 September 1951.

of movement is not an issue when the signal starts in the stele — the centre of a root — and moves outwards.

Questions for the future include: which sequences of amino acids in the SHR protein are required for its movement? What limits its movement to one cell layer and not more? And which other proteins participate? Moreover, the vasculature also seems to provide a positional signal in other developmental contexts, such as the differentiation of the surrounding photosynthetic cells in maize leaves¹⁰. Whether these events also rely on protein movement remains to be seen. ■

Sarah Hake is at the Plant Gene Expression Center, United States Department of Agriculture

Agricultural Research Service, and the University of California at Berkeley, 800 Buchanan Street, Albany, California 94710, USA.
e-mail: maizesh@nature.berkeley.edu

1. Nakajima, K., Sena, G., Naway, T. & Benfey, P. N. *Nature* **413**, 307–311 (2001).
2. van den Berg, C., Willemsen, V., Hage, W., Weisbeek, P. & Scheres, B. *Nature* **378**, 62–65 (1995).
3. Benfey, P. N. *et al. Development* **119**, 57–70 (1993).
4. Scheres, B. *et al. Development* **121**, 53–62 (1995).
5. Di Laurenzio, L. *et al. Cell* **86**, 423–433 (1996).
6. Wysocka-Diller, J. W., Helariutta, Y., Fukaki, H., Malamy, J. E. & Benfey, P. N. *Development* **127**, 595–603 (2000).
7. Helariutta, Y. *et al. Cell* **101**, 555–567 (2000).
8. Hake, S. & Char, B. *Genes Dev.* **11**, 1087–1097 (1997).
9. Sessions, A., Yanofsky, M. F. & Weigel, D. *Science* **289**, 779–781 (2000).
10. Langdale, J. A., Zelitch, I., Miller, E. & Nelson, T. *EMBO J.* **7**, 3643–3651 (1988).

Superconductivity

Up on the C₆₀ elevator

Paul Grant

The search for materials that lose electrical resistance — that is, become superconducting — at ever higher temperatures continues to pay dividends.

Organic materials do not naturally make good superconductors — those ‘perfect’ conductors that lose their electrical resistance completely when cooled below a certain transition temperature, T_C . But given the woefully low T_C values of most metallic superconductors, a lot of work has gone into mining more exotic materials, such as organic compounds, in the hope of finding higher- T_C nuggets. For most of my career, I was one of those diggers. So it is heartening to see Schön, Kloc and Batlogg hit pay dirt with the observation in *Science*¹ of superconductivity at an astounding 117 K in a special form of crystalline carbon. The temperature at which these C₆₀ crystals become superconducting is now high enough to begin challenging the existing record for superconductivity, which stands at 164 K and is held by the equally exotic copper oxide compounds.

The first organic materials found to exhibit superconductivity were a series of charge-transfer compounds, discovered in 1979 by Jérôme, Bechgaard and colleagues², with T_C values of around 1 K. This family, christened the Bechgaard salts, topped out at a T_C of 10–13 K and its superconducting properties were soon forgotten in the excitement that followed the discovery of copper oxide compounds with T_C values above 100 K.

Interest in organic superconductors then waned for half a decade until, in 1991, a T_C of around 18 K was discovered in C₆₀ crystals doped with alkali metals, an astonishingly high temperature for an organic-based substance³. Because crystalline C₆₀ is a natural insulator, it has to be chemically altered to become superconducting. By adding elec-

trons to the empty conduction band, or removing them from the valence band (creating positively charged holes), the crystals become conductors — and, at low enough temperatures, superconductors. This is what doping with alkali metals did.

In the past few years, Batlogg and colleagues have found another way to turn organic materials into superconductors^{4–7}. They incorporate the organic compound into the heart of an electronic device called a field-effect transistor (FET), which allows them to inject electrons or holes into the material simply by changing the voltage applied across the transistor. In the case of pure, unadulterated C₆₀, pumping it with electrons transforms it from an insulator into a superconductor with a transition temperature of only 10 K, but doping it with holes raises its transition temperature to a respectable 52 K.

The latest advance by Schön *et al.*¹ builds on their record of 52 K for hole-doped C₆₀ by modifying its structure with inert CHBr₃ and CHCl₃, molecules that expand the crystal lattice in much the same way that alkali metals do. The fact that these particular molecules are inert — they do not donate or accept any electrons — does not matter. The transition temperature of alkali-doped C₆₀ increases linearly with lattice spacing, and researchers expected a similar effect with hole-doped C₆₀, except that it was difficult to add holes to C₆₀ by chemical means. Schön and colleagues’ ability to inject holes using an FET device led to their latest round of experiments.

Here we need a short digression into the theory behind ‘conventional’ superconduct-

tivity, which includes C₆₀ but excludes the copper oxides (for which there is no consensus yet). Back in 1957, Bardeen, Cooper and Schrieffer⁸ (BCS) showed — in one of the great ironies of condensed-matter physics — that the tiny lattice vibrations (phonons) which cause normal electrical resistance, can, in certain circumstances, bind charge carriers (electrons or holes) into superconducting ‘pairs’ that flow without resistance.

In the BCS theory, the strength of this charge-phonon pairing, known as λ , is closely linked to the transition temperature, so the greater the value of λ , the higher the T_C . In turn, λ is the product of two other quantities: the number of charges per unit of energy available for pairing, known as the ‘density of states’, N_0 ; and the effect of the vibrations of the crystal lattice on the motion of these charges, or, in quantum-mechanical parlance, the ‘charge-phonon matrix element’, V . One way to increase λ , and thus T_C , is to try to increase N_0 by stretching out the lattice as much as possible so as to pack more mobile charges into a narrower energy range. There is some evidence — perhaps now mostly forgotten — that this effect occurs in the Bechgaard salts.

Schön *et al.*¹ now show that it is possible to ‘engineer’ the critical properties of C₆₀ in a way that is difficult to do in other structures. In most conventional superconductors, N_0 and V are intricately intertwined — in quantum-mechanical parlance, they are based on the same electronic wavefunctions — and it is hard to change N_0 through materials engineering without affecting V . But in solid C₆₀ crystals, V is determined by the vibration dynamics of the wavefunctions of the carbon atoms⁹, and N_0 by the standing wavefunctions between them. So C₆₀ offers a chance to structurally adjust N_0 independently of V .

And that is how Schön *et al.* did it. By expanding the C₆₀ lattice with CHBr₃, they increased N_0 with little or no apparent effect on V , and achieved a T_C of 117 K. In doing so, they demolished all notions that ‘conventional’ superconductors would be forever stuck below 50 K.

I have no idea what applications a superconducting C₆₀ FET might have. Attempts to use superconducting FETs as power switches have found the device currents to be too small. They are of the order of 1 ampere, whereas conventional semiconductor power can readily handle up to 2,000 amperes at 9,000 volts. The best hope for applications is somehow to reproduce the FET results in a bulk material that is several cubic centimetres in volume, rather than in the nanometre-thick conducting channel of an FET. But let us enjoy this advance in the science of superconductivity for its own sake. It is great to see that we can keep riding the rising elevator of high-temperature superconductivity in organic compounds.

How much higher can T_C go in C_{60} ? Schön and colleagues suggest that a further 1% increase in C_{60} separation might result in a T_C of around 150 K. (The existing all-time record for superconductivity is 164 K for copper oxides subjected to extreme pressures.) Any greater expansion of C_{60} would probably destroy its intermolecular bonding completely, but there might be other, more intriguing ways to take advantage of the unique crystal structure of this compound. The separation between the electronic states and vibrational modes of C_{60} exploited by Schön *et al.* is just the sort of thing Little had in mind in the 1960s when he proposed his 'excitonic' model of superconductivity¹⁰. According to this idea, superconducting paired carriers are glued together by excitons with characteristic energies, unlike the several hundred degrees of phonons, of around 20,000 K — potentially yielding

T_C values well above room temperature. Could new excitonic states be introduced by clever chemical or structural modification of C_{60} molecules? Don't step off the elevator just yet.

Paul Grant is at the Electric Power Research Institute, 3412 Hillview Avenue, Palo Alto, California 94303, USA.

e-mail: pgrant@epri.com

- Schön, J. H., Kloc, Ch. & Batlogg, B. *Science*, 30 August 2001 (10.1126/science.1064773).
- Jérome, D. *et al.* *J. Phys. Lett. (Paris)* **41**, L95 (1980).
- Hebard, E. F. *et al.* *Nature* **350**, 600–601 (1991).
- Schön, J. H., Kloc, Ch. & Batlogg, B. *Nature* **406**, 702–704 (2000).
- Schön, J. H., Kloc, Ch. & Batlogg, B. *Nature* **408**, 549–552 (2000).
- Schön, J. H. *et al.* *Nature* **410**, 189–192 (2001).
- Schön, J. H., Kloc, Ch., Hwang, H. Y. & Batlogg, B. *Science* **292**, 252–254 (2001).
- Bardeen, J., Cooper, L. N. & Schrieffer, J. R. *Phys. Rev.* **108**, 1175–1204 (1957).
- Ramirez, A. P. *et al.* *Phys. Rev. Lett.* **68**, 1058–1060 (1992).
- Little, W. A. *Sci. Am.* **212**, 21 (1965).

Ecology

Crowd trouble for predators

Peter D. Moore

Flocking, herding, swarming: call it what you will. But when you're somebody's lunch there's safety in numbers, even when the predator is an aquatic plant.

Avoiding being eaten is (obviously) one way that most organisms escape premature death. There are many techniques for evading predation, one of the most frequent being the herding or flocking

behaviour often seen in certain mammals, birds and fish. But what if the predator happens to be a plant? Does it still help to gather together in groups for mutual support? From work reported by Goran Englund and Sabine Harms (*Oikos* **94**, 175–181; 2001), it seems that it does. Their finding in turn raises such questions as how the prey species (in this case, microscopic aquatic creatures called cladocerans) know when to swarm, and how the swarming mechanism operates.

The plant concerned, the bladderwort *Utricularia vulgaris* (Fig. 1), traps incautious zooplankton within its hollow leaves where it digests them and absorbs their nutrients. Although it is a sedentary predator, bladderwort has an active trapping mechanism, so its carnivory is not entirely passive. Sensory hairs on the outside of the bladder-like leaf detect the local presence of potential prey, resulting in the opening of a trap door and a rush of water and prey into the hollow leaf as negative pressure is released. But being static, the plant must rely on the errant prey organism colliding with the trap. One might predict, therefore, that the higher the density of the zooplankton, the more likely such collisions would become and the more successful the predator would be. Englund and Harms, however, show that at high densities

of cladoceran the number of prey trapped per bladder actually declines.

Observations of the behaviour of the prey, *Polyphemus pediculus*, revealed that they are either randomly or uniformly dispersed at low densities (20 individuals in a 125-ml vessel), whereas at high densities (120 individuals per vessel) they aggregate into swarms. What constitutes the immediate stimulus for this response (such as visual or chemical signals) is unclear, but the flocking behaviour may well be the factor responsible for the reduced success of the predator. Two main changes occur in the swimming behaviour of individual cladocerans when they occur in a swarm — they swim more slowly, and they follow a more contorted path than when they are dispersed.

Obviously, a slower-moving animal is less likely to encounter a static object within a given time, so the chances of a cladoceran hitting a trap are reduced if it slows down. It is also possible that at slower speeds the animal is better able to identify and avoid the traps. Again, it could be argued that a linear motion on the part of the prey is more likely to lead it to traps than random motion. However, given the sensory capacity of the *Polyphemus*, motion may not actually be random, but selective to avoid potentially dangerous objects, such as *Utricularia* leaves.

Another question that requires investigation is the possibility of predator satiation. If the plant has simply had its fill when its prey is abundant, then further increases in prey density would not increase trap success. But in these experiments trap success actually declines as density increases (over a threshold of about 35 animals per 125-ml vessel). Also, the 'handling rate' for a trap — the time it takes for the trapping mechanism to be reset — is around 10–15 minutes, so satiation is not a likely explanation of the observations.

It is of course possible that the swarming behaviour of *Polyphemus* is primarily the outcome of some entirely different behavioural demand. Perhaps it provides better feeding opportunities, or is related to reproductive cycles, or results in more efficient escape from mobile predators. Whatever the reason, swarming still gives the animal a better chance of avoiding consumption by a predatory plant.

Sticking together has proved an effective defence mechanism for many animals subjected to predation, from wildebeest to mackerel. The fact that it works even among microscopic creatures threatened by enemies as passive as a plant may provide insights into the benefits of the swarm.

Peter D. Moore is in the Division of Life Sciences, King's College, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NN, UK.
e-mail: peter.moore@kcl.ac.uk

Figure 1 Bladderwort — armed and dangerous (at least to cladocerans).



Nonlinear physics

Déjà vu in optics

Nail N. Akhmediev

Using optical fibres, experimentalists have confirmed that a physical version of *déjà vu* — whereby a system returns to its original state — does occur with light waves.

Have you ever experienced a phenomenon called *déjà vu*? An event from the past appears to be happening again as clearly as the first time you experienced it. It is a neurological phenomenon whose roots are mysterious and unclear. No matter how complicated the processes in our brains may be, *déjà vu* can have analogues in the inanimate world. In physics it is called the Fermi–Pasta–Ulam recurrence. As reported in *Physical Review Letters*, Van Simaëys *et al.*¹ have now observed this recurrence phenomenon in a fibre-optics experiment.

Since the advent of computers, physicists have been using them to model the physical world. In 1953, the well-known physicist Enrico Fermi, together with two experts in mathematics, John Pasta and Stanislaw Ulam, used one of the earliest digital computers — a machine at Los Alamos called MANIAC — to model the properties of complex nonlinear systems². In particular, they wanted to simulate oscillations of a string consisting of 64 particles with nonlinear forces between them — such forces are nonlinear if equal amounts of the force do not always produce equal effects. Their intention was to study the ‘thermalization’ of the system — or how energy is distributed among the many possible oscillations generated by the string.

The results of the calculations were not what Fermi and his collaborators expected². After starting the system with all the energy

associated with one type of string oscillation (the initial mode), they expected that the total energy would gradually become shared between all the modes of the system. To their great surprise, the opposite happened. After a relatively short time, the energy returned to the same mode that was excited initially. To understand why this is so unexpected, consider this rough analogy: if we heat a piece of iron from one end, the heat will spread across the whole metal homogeneously. Clearly, no-one would expect that it would ever return to its original state — with only one end being hot.

Some ten years later, in an attempt to solve this mystery, Zabusky and Kruskal³ used more advanced computers and modified the equations used by Fermi, Pasta and Ulam. Specifically, they solved the Korteweg–de Vries equation, which describes shallow water waves or long-period waves in nonlinear crystal lattices. Using this approach to the Fermi–Pasta–Ulam problem, Zabusky and Kruskal numerically discovered in the system a wave which they called the ‘soliton’. This is a solitary wave that propagates without ever changing shape, and was first observed over 150 years ago by John Scott Russell as a hump of fast-moving water on the Edinburgh to Glasgow canal. Zabusky and Kruskal’s theoretical discovery of the soliton had a lasting impact on twentieth-century mathematical physics. Their work led to a number of important developments

related to ‘integrable nonlinear models’, in which solitons are one of the possible solutions. But the Fermi–Pasta–Ulam mystery was not completely solved. Although Zabusky and Kruskal observed a return to the approximate initial conditions, this was not a recurrence in the way that we understand this phenomenon today.

Later, with the development of mathematical physics, the nonlinear Schrödinger wave equation came to the fore as another completely integrable nonlinear equation that also has soliton solutions⁴. But it is not only solitons that are of interest here. It was previously shown^{5,6} that continuous waves described by the nonlinear Schrödinger equation will break up into small wave packets when there are slight perturbations (secondary waves) on top of the initial continuous wave. This process is known as a modulation instability and it was later found to be closely related to the problem of Fermi, Pasta and Ulam. The crucial point is that, when the secondary perturbation is regular — say it consists of a single mode — then after growing to some maximal amplitude, the perturbation reduces in size and eventually disappears. Like the letters in a palindrome, the initial unperturbed continuous wave is completely restored in reverse order⁷. So essentially the Fermi–Pasta–Ulam recurrence is a periodic solution of the nonlinear Schrödinger wave equation; it has a variable period (though the period may be infinitely long), and the theory has been developed exactly for this situation^{7,8}.

The Fermi–Pasta–Ulam recurrence has been much harder to demonstrate experimentally, even though the theory is known to be fairly accurate for a physical system such as an optical fibre. Figure 1 shows the theoretical evolution of the frequency components of the wave during the modulation instability of a continuous wave. Initially all the energy is concentrated in the central mode but, over time, energy flows from the central mode into the two weak seed modes. As a result, the energy in the central mode decreases. An infinite number of modes can become excited (although only three are shown in Fig. 1) until a certain point is reached at which the energy starts to flow back from the side modes into the central component. This reciprocal energy exchange between the central mode and an infinite number of side modes is the Fermi–Pasta–Ulam recurrence⁹.

In fact, it has taken another 16 years to observe this optical recurrence experimentally. In their experiment, Van Simaëys *et al.*¹ saw recurrence for the first time when they were observing the effect of a modulation instability in an optical fibre. By pumping energy into the central frequency mode at the input to an optical fibre, they were able to watch the energy spread between the frequency components, followed by all the

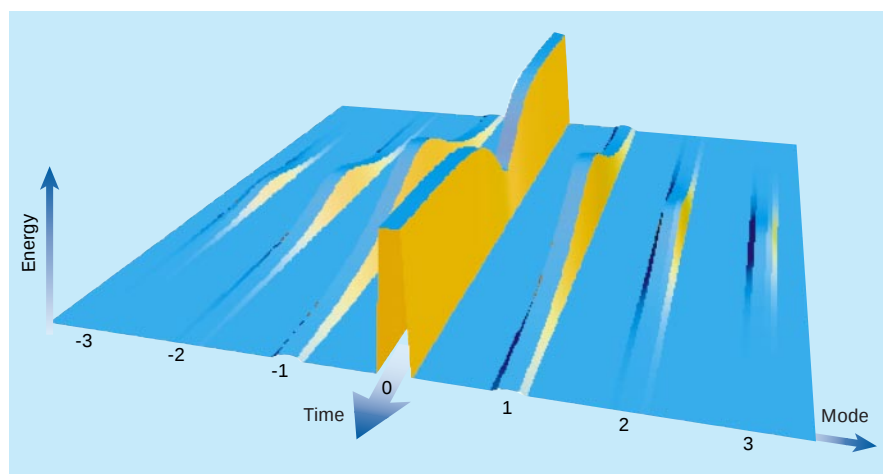


Figure 1 The Fermi–Pasta–Ulam recurrence as seen in this solution of the nonlinear Schrödinger wave equation. As the frequency components of a continuous wave evolve, energy flows from the central mode into the side modes. The energy eventually returns, restoring the wave to its initial state. Such a Fermi–Pasta–Ulam recurrence has been demonstrated by Van Simaëys *et al.*¹ in a recent fibre-optics experiment.

energy flowing back to the original mode, just as Fig. 1 shows. In other words, by sending a slightly modulated 'continuous' wave (actually a well-shaped rectangular pulse) along the optical fibre, the researchers were able to see an increase in the amplitude of the modulated components and subsequent restoration of the initial wave. Because they took great care when setting up the initial conditions, the recurrence they saw was almost perfect.

Nail N. Akhmediev is in the Optical Sciences Centre, Research School of Physical Sciences and Engineering, Australian National University, ACT 0200, Australia.

e-mail: nna124@rsphysse.anu.edu.au

1. Van Smaeyens, G., Emplit, Ph. & Haelterman, M. *Phys. Rev. Lett.* **87**, 033902 (2001).
2. Fermi, E., Pasta, J. & Ulam, S. *Studies of Nonlinear Problems I* (Los Alamos report LA-1940, 1955); reprinted in *E. Fermi. Collected Papers Vol II* (ed. Segrè, E.) 978 (Univ. Chicago Press, 1965).
3. Zabusky, N. J. & Kruskal, M. D. *Phys. Rev. Lett.* **15**, 240–243 (1965).
4. Zakharov, V. E. & Shabat, A. B. *Sov. Phys. JETP* **34**, 62–69 (1971).
5. Benjamin, T. B. & Feir, J. E. *J. Fluid Mech.* **27**, 417–430 (1967).
6. Bespalov, V. I. & Talanov, V. I. *Sov. Phys. JETP Lett.* **3**, 471–473 (1966).
7. Akhmediev, N. N. & Korreev, V. I. *Theor. Math. Phys.* **69**, 1089–1092 (1986).
8. Akhmediev, N. N. & Ankiewicz, A. *Solitons, Nonlinear Pulses and Beams* (Chapman & Hall, London, 1997).
9. Akhmediev, N. N., Heatley, D. R., Stegeman, G. I. & Wright, E. M. *Phys. Rev. Lett.* **65**, 1423–1426 (1990).

Cell cycle

Archipelago of destruction

Michael Schwab and Mike Tyers

The finely balanced activity of enzymes and their regulators keeps the cell-division cycle under control. A newly discovered molecule that ensures the timely destruction of one regulator is mutated in some cancer cells.

The cell cycle is the process by which cells grow, replicate their genome, segregate the two copies of the genome, and divide. The progression of these events is controlled by enzymes known as cyclin-dependent kinases (CDKs), whose activity in turn depends on their association with proteins called cyclins. For example, Cdk2 works with cyclin E to catalyse two key processes — the start of DNA replication and duplication of the centrosome, the precursor to the chromosome-separating apparatus¹. The onset and demise of CDK activity is controlled mainly by the scheduled degradation of proteins that inhibit CDKs and of cyclins². The timely appearance and disappearance of cyclin E is crucial: excessive activity of the cyclin E–Cdk2 complex drives cells to copy their DNA prematurely, resulting in genome instability³; moreover, cyclin E levels are often higher than normal in human tumours⁴. Three groups, writing on pages 311 and 316 of this issue^{5,6} and in *Science*⁷, have now identified the crucial factor that directs cyclin E to the degradation machinery.

The programmed degradation of many proteins that regulate the cell cycle is carried out by the ubiquitin–proteasome system. A cascade of enzymes⁸ — generically termed E1, E2 and E3 — catalyses the addition of polymers of the small protein ubiquitin to the protein substrates. This polyubiquitin tag acts as an address label, directing substrates to the 26S proteasome, where they are destroyed. Specificity in the ubiquitin system is controlled largely by E3 enzymes, also known as ubiquitin ligases.

Ubiquitin ligases termed SCF complexes are necessary for the onset of S phase (the

cell-cycle stage during which DNA is replicated)². SCF complexes consist of three core subunits, which couple to any one of several 'F-box' proteins (Fig. 1) — so called because they contain the F-box, a sequence of amino acids first identified in cyclin F⁹. The F-box proteins also contain certain domains, such as the WD40 motif, that recognize specific substrates, usually ones that have been phosphorylated. Genetic evidence^{10,11} has implicated several proteins in the destruction of cyclin E, including an F-box protein called Skp2. However, the presumed F-box protein that controls the specificity of ubiquitination of phosphorylated cyclin E has, until now, escaped identification.

As is often the case, the breakthrough came from studies of model organisms — in this case, by investigating the genes needed for cyclin E degradation. Moberg *et al.*⁵ started by screening fruitflies (*Drosophila melanogaster*) for mutant cells that have a proliferative advantage over normal cells. From this screen, they picked out the *archipelago* (*ago*) gene, and determined that levels of cyclin E protein — but not cyclin E messenger RNA — are higher than normal in *ago* mutant cells. These and other results imply that the degradation of cyclin E protein does not proceed normally in these cells, and that *ago* is involved in cyclin E degradation. The *ago* gene encodes a counterpart of the budding yeast Cdc4 and nematode and mouse SEL-10 proteins — related F-box proteins implicated in the degradation of various substrates.

Meanwhile, Strohmaier *et al.*⁶ and Koepf *et al.*⁷ used yeast cells with mutations in various components of the SCF complex as an entry point into the cyclin E degradation

pathway. Both groups found that cyclin E is stabilized in cells with mutations in Cdc4, and that the human counterpart of Cdc4 — which they respectively term hCdc4 and Fbw7 — specifically interacts with cyclin E. This interaction depends on cyclin E being phosphorylated on the threonine at position 380 in the amino-acid chain, an event also required for cyclin E degradation *in vivo*¹⁰. Mutations in the WD40 domain of hCdc4/Fbw7 stop it from binding to phosphorylated cyclin E, implying that this domain forms a docking site for phosphorylated substrates. Gratifyingly, a fully assembled SCF complex containing hCdc4/Fbw7 robustly ubiquitinates phosphorylated cyclin E *in vitro*.

In some tumours, cyclin E protein often accumulates at high levels⁴. So an obvious question was whether mutant forms of hCdc4/Fbw7/Ago occur in cancer cells. Moberg *et al.*⁵ and Strohmaier *et al.*⁶ show that several tumour cell lines that have high levels of cyclin E protein also bear mutations in the WD40 domain of hCdc4/Fbw7/Ago. This and other evidence^{5,6}, including a phenomenon known as loss of heterozygosity, seen in two cell lines, suggests that hCdc4/Fbw7/Ago is a tumour suppressor — a negative regulator of cell proliferation that normally prevents cancer progression. In multicellular organisms, cells with mutations in this protein that lead to a reduction in its function, and so to high levels of cyclin E protein, might gain a proliferative advantage over normal cells. More important, such mutations might also result in genome instability, the driving force behind the changeable characteristics of early cancer cells³.

In a second compelling disease connection, it turns out that SEL-10 (the nematode and mouse counterpart of hCdc4/Fbw7/Ago) is implicated in the control of signalling pathways involving the presenilin or Notch proteins^{12–14}. Alterations in the presenilin signalling pathway are thought to underlie some forms of Alzheimer's disease, so it will be important to find out whether hCdc4/Fbw7/Ago is involved in this pathway. And some activated forms of Notch are potent oncogenes, promoting cell division¹⁵, so disruption of hCdc4/Fbw7/Ago may well deregulate cell division at more than one level. It should be possible to test these ideas by deleting the corresponding gene in mice.

Some mechanistic details of the degradation of cyclin E have yet to be unravelled. For example, what is the role of the F-box protein Skp2 — a long-time contender for involvement in cyclin E degradation? Although cyclin E is stabilized in mouse cells that lack Skp2 (ref. 11), Strohmaier *et al.*⁶ found that, in side-by-side *in vitro* comparisons with hCdc4/Fbw7/Ago, Skp2 neither binds nor ubiquitinates cyclin E. So the effects of Skp2 on cyclin E might be indirect. Skp2 targets the Cdk2 inhibitor p27 for ubiquitination

Daedalus

Genomes and souls

The evolution of humanity is still a theological problem. Even the Catholic Church accepts evolution, but excludes the human soul from it. So, says Daedalus, presumably there was an advanced ape without a soul who gave birth to a baby that had one. Some mutation had created the human soul, our spiritual nature.

All the standard queries of evolution then cry out to be asked. If the soul is recessive, a creature with a soul mating with one with no soul would give soulless offspring. If it is dominant, the soul would spread rapidly, and (as claimed) all humanity would soon have souls. Furthermore, the argument implies that the soul is represented on the genome. One or more of the tens of thousands of genes in the human genome codes not for a protein, but for the soul.

Now that the human genome has been roughly described in its entirety, DREADCO biochemists are scanning it for a set of unusual bases conspicuously absent from the genomes of the higher animals. With luck they will identify such a set, associated with the soul. The way will then be open for several experiments, all highly controversial. First, this set of bases could be inserted in plasmid form into organisms too primitive to have much use for a soul, such as bacteria of the species *Escherichia coli*. Would they behave any differently from normal organisms of these species? Probably not; Daedalus reckons that a brain of some complexity is needed to support a soul. Either the transplant would fail, or the bases could not be switched on.

Alan Turing once suggested that a creature with a brain arguably capable of supporting a soul, such as an elephant, might be given one. In this case, new behaviour (such as praying) should certainly occur. Second, human beings deprived of the crucial bases, or perhaps with them 'switched off' by some binding ligand, should react in an inhumanly 'soulless' or destructive manner. But if the Lord objected to interference with His creation, the attempt to deprive humans of their souls would somehow fail.

And, of course, the genome occurs in every cell of the body. So transplant operations, in which organs are transferred from one person to another, would produce a chimaera with two souls, at least in theory. This suggests that the ultimate surgical operation, a brain transplant, will never work. It also suggests that no computer, however clever, can have a soul.

David Jones

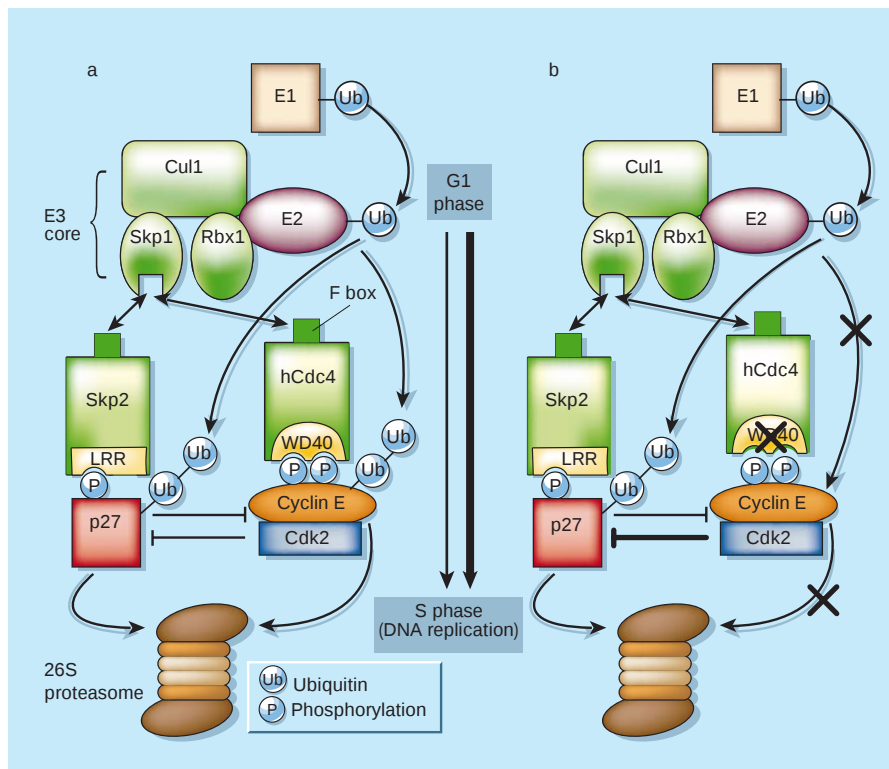


Figure 1 Cell-cycle balancing acts. The enzyme Cdk2 is activated only at the transition from the G1 phase to the S phase of the cell cycle. This is achieved by a complex interplay between degradation of cyclin E, which is needed for Cdk2 to function, and of the Cdk2 inhibitor p27. Degradation occurs as follows. E1, E2 and E3 (green) enzymes add ubiquitin to substrate proteins, which are then directed to the 26S proteasome for destruction. New work⁵⁻⁷ shows that hCdc4/Fbw7/Ago (shown as hCdc4) links phosphorylated cyclin E to a core E3 complex, termed SCF. Similarly, Skp2 recognizes phosphorylated p27 (ref. 16). **a.** In normal cells¹, growth factors stimulate the expression of both Skp2 and cyclin E, which cooperate to ensure the degradation of p27. With precise timing, the balance of activity is tipped in favour of cyclin E–Cdk2, which promotes S-phase entry. **b.** In tumour cells, mutation of Cdc4 prevents recognition of cyclin E, which is stabilized. This overwhelms p27, resulting in premature S-phase entry. WD40 and LRR are protein regions that recruit substrates for ubiquitination.

and degradation, but only once the inhibitor has been phosphorylated by cyclin E–Cdk2 (ref. 16; Fig. 1). So cells lacking Skp2 would have excess p27, which would inhibit Cdk2 and prevent it from phosphorylating cyclin E. Cyclin E would not be recognized by hCdc4/Fbw7/Ago, and so would be stabilized. Alternatively, ubiquitination of cyclin E by Skp2 might require other unknown proteins that were not included in the *in vitro* experiments.

A second issue concerns the phosphorylation requirements for cyclin E recognition. In yeast, many substrates of Cdc4 must be phosphorylated on several sites before they can be recognized. Similarly, Strohmaier *et al.* find that the recognition of cyclin E depends on it being phosphorylated on at least two residues — threonines 62 and 380. This observation raises the possibility of several intriguing regulatory mechanisms, such as the existence of kinases in addition to Cdk2 that might target cyclin E for degradation. A detailed understanding of how hCdc4/Fbw7/Ago functions will be essential if the new results⁵⁻⁷ are to be exploited in treating cancer and other diseases.

Michael Schwab and Mike Tyers are in the Programme in Molecular Biology and Cancer, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto M5G 1X5, Canada.

e-mail: tyers@mshri.on.ca

- Ekholm, S. V. & Reed, S. I. *Curr. Opin. Cell Biol.* **12**, 676–684 (2000).
- Zachariae, W. & Nasmyth, K. *Genes Dev.* **13**, 2039–2058 (1999).
- Spruck, C. H., Won, K. A. & Reed, S. I. *Nature* **401**, 297–300 (1999).
- Keyomarsi, K. & Herliczek, T. W. *Prog. Cell Cycle Res.* **3**, 171–191 (1997).
- Moberg, K. H., Bell, D. W., Wahrer, D. C. R., Haber, D. A. & Hariharan, I. K. *Nature* **413**, 311–316 (2001).
- Strohmaier, H. *et al.* *Nature* **413**, 316–322 (2001).
- Koepp, D. M. *et al.* *Science*, 30 August 2001 (10.1126/science.1065203).
- Hershko, A. & Ciechanover, A. *Annu. Rev. Biochem.* **67**, 425–479 (1998).
- Bai, C. *et al.* *Cell* **86**, 263–274 (1996).
- Winston, J. T., Chu, C. & Harper, J. W. *Genes Dev.* **13**, 2751–2757 (1999).
- Nakayama, K. *et al.* *EMBO J.* **19**, 2069–2081 (2000).
- Hubbard, E. J., Wu, G., Kitajewski, J. & Greenwald, I. *Genes Dev.* **11**, 3182–3193 (1997).
- Öberg, C. *et al.* *J. Biol. Chem.*, 18 July 2001 (10.1074/jbc.m103922001).
- Wu, G., Hubbard, E. J., Kitajewski, J. K. & Greenwald, I. *Proc. Natl Acad. Sci. USA* **95**, 15787–15791 (1998).
- Ellisen, L. W. *et al.* *Cell* **66**, 649–661 (1991).
- Carrano, A. C., Eytan, E., Hershko, A. & Pagano, M. *Nature Cell Biol.* **1**, 193–199 (1999).

Obituary

Fred Hoyle (1915–2001)

Fred Hoyle, who died on 20 August at the age of 86, was the most imaginative of men, a kind of Leonardo. He made monumental contributions to astrophysics and cosmology, and was a brilliant popularizer (and science-fiction writer and occasional playwright). He also put his name to much rubbish and was embroiled in controversy for most of his life.

Hoyle was born and brought up in Yorkshire and remained a true son of that county — directly spoken, combative, careful with money and self-reliant. He learned to read, he said, from subtitles on the silent films he saw at the cinema, while playing truant from school. How natural, then, that he should become a scholarship boy at the University of Cambridge, in 1933. And that over the following three years he should work his way as a mathematician to the top of his class.

The ensuing period as a research student was the best of Hoyle's professional life. With Rudolf Peierls as his supervisor, he had an admission ticket to the weekly theoretical seminar at the Cavendish Laboratory attended by the luminaries of the day. Hoyle then persuaded Paul Dirac to take him on as a student (by promising Dirac that supervision would not be required), and in due course produced an item of world-class thinking. With Ray Lyttleton he showed how to calculate the luminosity of a star from knowledge of its mass, so providing a theoretical basis for Arthur Eddington's empirically derived relationship between stellar mass and luminosity.

After work on radar during the Second World War, Hoyle's career had three main strands: the constitution of stars (made tractable by Hans Bethe's 1937 account of how stars can get their energy by the conversion of hydrogen via deuterium to helium); nucleosynthesis (the production in stars of elements heavier than helium); and 'steady-state' cosmology (the idea, to be contrasted with Big Bang theory, that the Universe is the same everywhere and its expansion can be accounted for by the continual creation of new matter).

The first theme led directly to Hoyle's founding in 1966 of the Institute of Theoretical Astronomy at Cambridge, with the self-mocking acronym IOTA. The institute's foundation came seven years after it had been proposed, and only after bruising rows with the university and the UK Science Research Council. It was subsequently renamed the Institute of Astronomy.



Cosmologist and controversialist

Nucleosynthesis had a happier outcome. Hoyle had won his spurs by predicting that carbon can be synthesized plentifully in stars (as it is) only if there is a metastable excited state of the carbon nucleus; W. A. Fowler, with his van de Graaff machines, showed that to be correct. Hoyle and Fowler then accounted for the elements up to iron, and Margaret and Geoffrey Burbidge, with Fowler's collaboration, sought to make sense of the abundances of the heavier elements. Finally, all four put their heads together and in 1957 published in *Reviews of Modern Physics* the classic paper now known affectionately as B²FH. Fowler won a Nobel prize for his work. Hoyle, shamefully, did not.

The business of the steady-state Universe has always been more contentious. The originators of the idea were Hermann Bondi and Thomas Gold; Hoyle was an interloper of a kind. Perhaps he relished the inevitable row. But he did not enjoy the personal animosity engendered between him and the late Martin Ryle, then one of the big-hitters in British astronomy. Hoyle believed that the ill-feeling went back to a disagreement at a conference in 1951, in which he had come out on top. Ten years later Ryle got his own back at a media event, which — so general opinion had it — spelled the end of the steady-state theory with Hoyle's seeming failure to respond to the evidence marshalled against it.

None of this had prevented Hoyle from being elected to Eddington's old chair in 1958 and, later, being chosen as chairman of the Science Research Council's main

astronomy committee. In that role, he was influential in putting the Anglo-Australian Telescope on a sound administrative footing, and in the mid-1960s accepted what proved to be a poisoned chalice — chairing a review of telescope provision in the Northern Hemisphere for British astronomers.

Initially, Hoyle was as cheerful as I ever knew him: he bubbled with conviction that much could be done at reasonable cost. But the cards were stacked against him, and the upshot of long-drawn-out manoeuvring was a stitch-up by his opponents. Following defeat at a crucial meeting in November 1971, Hoyle quit his chair at Cambridge.

Then what? Without a salary and five years short of pensionable age, Hoyle had to earn money, and quickly. So he spent two years lecturing in the United States before returning to Britain and diving into controversy from his position, assumed in 1975, as visiting professor at the University of Cardiff. This was a post sponsored by Chandra Wickramasinghe, with whom Hoyle published his most controversial views. The most egregious example was the allegation that the British Natural History Museum's specimen of *Archaeopteryx* is a forgery. Even his best friends were at a loss to know why he behaved in such a way in this and other instances.

But Hoyle was by no means off his head. In 1994, with Jayant V. Narlikar he published in *Reviews of Modern Physics* a meticulous account of how the current formalism of quantum field theory can be replaced by the formalism of advanced and retarded potentials that sprang from Maxwell's time. If ever quantum-field theorists want to return to properly relativistic action-at-a-distance, there is their recipe. And only last year came the book *A Different Approach to Cosmology*, by Hoyle, Geoffrey Burbidge and Narlikar, published by Cambridge University Press. This is a scholarly rather than a polemical work, in that it is a well-documented guide to extragalactic evidence against the Big Bang (a term, incidentally, intended to be derisive, that was coined by Hoyle himself in a broadcast lecture of 1952). My own conviction is that Hoyle's scepticism was well-founded. But it is too soon to tell how the Big Bang will be replaced by some other cosmology. That is heterodoxy. Soon after *A Different Approach to Cosmology* appeared, I asked one of Hoyle's former colleagues if he'd read it. "Wouldn't waste the time," was the reply. The hope must be that civility will break out among Hoyle's enemies now that he is dead.

John Maddox

John Maddox is emeritus editor of *Nature*.

Neonatal sunburn and melanoma in mice

Severe sunburn in newborn, but not adult, mice is linked with melanoma in later life.

Retrospective epidemiological data have indicated that cutaneous malignant melanoma may arise as a consequence of intense, intermittent exposure of the skin to ultraviolet radiation, particularly in children, rather than from the cumulative lifetime exposure that is associated with other forms of skin cancer^{1–3}. Here we use a genetically engineered mouse model to show that a single dose of burning ultraviolet radiation to neonates, but not adults, is necessary and sufficient to induce tumours with high penetrance which are reminiscent of human melanoma. Our results provide experimental support for epidemiological evidence that childhood sunburn poses a significant risk of developing this potentially fatal disease.

Although transgenic mice have been used to describe a few cutaneous melanoma models (reviewed in ref. 4), murine melanocytic tumours are dermal in origin and lack the epidermal component that characterizes conventional human melanoma — probably because murine melanocytes, unlike those in human skin, are confined to hair follicles. However, the skin of transgenic mice in which a metallothionein-gene promoter forces the overexpression of hepatocyte growth factor/scatter factor (HGF/SF) has melanocytes in the dermis, epidermis and dermal–epidermal junction, and is thus more akin to human skin^{5,6}.

HGF/SF, the ligand of the receptor tyrosine kinase c-Met, is a multifunctional regulator of cellular growth, motility and invasiveness; aberrant c-Met signalling has been implicated in human cancer, including cutaneous melanoma^{7–9}. Aged HGF/SF-transgenic mice develop sporadic melanoma with metastasis^{5,6} without exposure to ultraviolet light, whereas the formation of non-melanocytic cutaneous neoplasms, but not melanoma, is accelerated by chronic, suberythral (insufficient to redden skin) ultraviolet irradiation of adult mice¹⁰.

We subjected albino HGF/SF-transgenic mice and wild-type littermates to erythral (skin-reddening) ultraviolet irradiation at 3.5 days of age, 6 weeks of age, or both. We found that a single neonatal dose (group C), which was 30-fold lower than the total ultraviolet dose administered previously to adult mice¹⁰, was sufficient to induce melanoma in HGF/SF-transgenic mice after a relatively short latent period and with high cumulative incidence (Fig. 1a). This neonatal dose roughly corresponds to a sunburning dose of natural sunlight at midlatitudes in midsummer (23 standard erythral doses; see supplementary information).

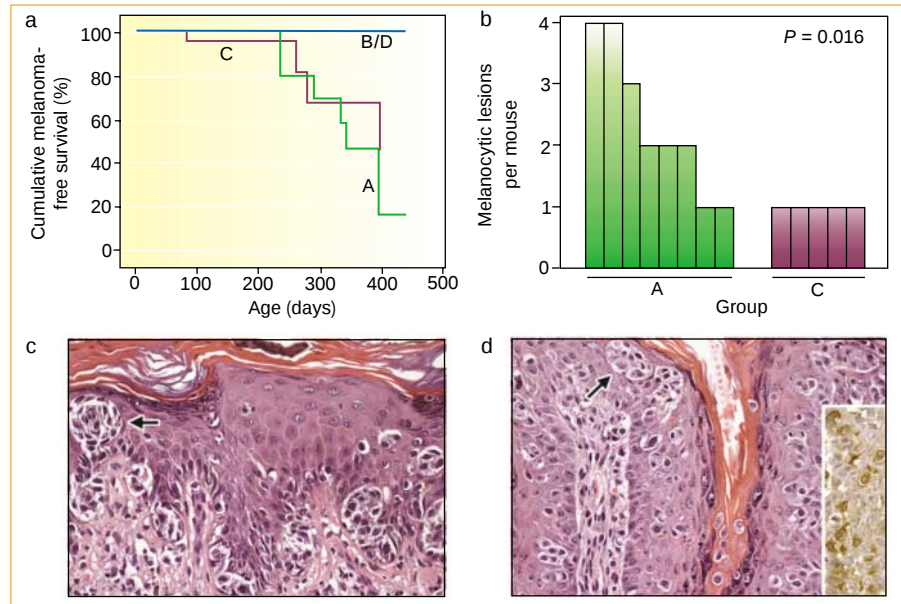


Figure 1 HGF/SF-transgenic mice develop tumours reminiscent of cutaneous malignant melanoma in response to neonatal erythral ultraviolet (UV) irradiation. **a**, Cumulative survival of melanoma-free transgenic mice as a function of age (Kaplan–Meier analysis) in cohorts UV-irradiated at: group A, 3.5 days and again at 6 weeks; group B, 6 weeks; group C, 3.5 days; group D, no UV treatment. Mice were generated on an albino inbred FVB background. Neonatal mice received a single 15-min treatment with Phillips F40 UV lamps for a total unweighted UV dose^{12,13} of 9.58 kJ m^{-2} (UV-A, 320–400 nm, 3.31 kJ m^{-2} ; UV-B, 280–320 nm, 6.24 kJ m^{-2} ; UV-C, 250–280 nm, 0.03 kJ m^{-2} (UV-C is not present in terrestrial sunlight but is not essential for development of melanocytic lesions in this mouse model; our unpublished results)). Neonatal UV treatment was erythemogenic and caused occasional desquamation but no mortality. Six-week-old mice received a single treatment of 19.16 kJ m^{-2} . **b**, Multiplicity of melanocytic lesions for individual transgenic animals in groups A and C; the two groups were significantly different (Mann–Whitney U-test). **c**, Human microinvasive melanoma exhibiting solitary and nested atypical melanocytes along and above the dermal–epidermal junction, so-called ‘pagetoid spreading’ (arrow shows a nest of melanocytes). **d**, Comparable lesion in a UV-irradiated HGF/SF-transgenic mouse (arrow shows a nest of melanocytes). Inset, melanocyte-specific immunohistochemical staining of tyrosinase-related protein-1. For further details, see supplementary information.

Melanoma development in HGF/SF-transgenic mice after ultraviolet irradiation at both 3.5 days and 6 weeks (group A, Fig. 1a) was indistinguishable from that seen after only a single exposure at 3.5 days (group C), whereas a similar dose at 6 weeks (group B) was not tumorigenic. However, the second exposure to ultraviolet light increased the multiplicity of melanocytic lesions (Fig. 1b), as well as the incidence of non-melanocytic tumours, including squamous-cell carcinoma and sarcoma (see supplementary information). Melanomas were not seen in either non-transgenic or untreated transgenic mice during the course of the experiment (13 months).

In contrast to other mouse melanoma models, most melanomas (9 of 11) that arose in our HGF/SF-transgenic mice as a result of neonatal ultraviolet irradiation were highly interactive with the epidermis (that is, they showed junctional activity), and resembled lesions in human patients at the various stages of progression, including metastasis (Fig. 1c, d). Molecular analysis indicates that, as in human melanoma⁷,

these murine tumours acquire receptor tyrosine kinase autocrine signalling loops (HGF/SF–Met), frequently lose the *ink4a* gene, and rarely contain p53 mutations (results not shown).

Why is neonatal mouse skin so sensitive to ultraviolet-induced melanomagenesis? As melanocyte-progenitor cells are more abundant in neonatal than in adult skin and more proliferative under stress¹¹, ultraviolet exposure may stimulate proliferation of DNA-damaged neonatal progenitors and thus facilitate melanomagenesis. Early exposure to intense ultraviolet radiation might also affect the developing immune system, promoting future tolerance to arising melanoma.

Caution is called for in extrapolating these results to sunburn in children. The difference in thickness between mouse and human skin could affect the penetration of ultraviolet radiation, although UV-A and UV-B light of longer wavelengths will penetrate to the epidermis and dermis in human as well as murine skin. Moreover, the human age equivalent to a 3.5-day-old

neonatal mouse cannot be precisely calculated. Untreated HGF/SF-transgenic mice are already genetically predisposed to late-onset melanoma^{5,6}, a risk situation that more closely resembles that in *INK4a*-deficient, melanoma-prone kindreds than in the general human population.

However, our experimental mouse model resembles human cutaneous malignant melanoma with respect to aetiology, histopathology and molecular pathogenesis. It is consistent not only with a causal link between childhood sunburn and the development of melanocytic neoplasms, but also with possible promotion by subsequent ultraviolet exposure. The HGF/SF-transgenic mouse should help in the assessment of genetic and environmental risk factors for melanoma, as well as in developing protection against sunburn and melanomagenesis.

Frances P. Noonan*, **Juan A. Recio†**,
Hisashi Takayama†, **Paul Duray†**,
Miriam R. Anver§, **Walter L. Rush||**,
Edward C. De Fabo*, **Glenn Merlino†**

*Laboratory of Photobiology and
Photoimmunology, Departments of Dermatology
and Immunology, George Washington University
Medical School, Washington DC 20037, USA

†Laboratory of Molecular Biology, National Cancer
Institute, Bethesda, Maryland 20892-4255, USA
e-mail: gmerlino@helix.nih.gov

‡Laboratory of Pathology, National Cancer
Institute, Bethesda, Maryland 20892, USA
§Pathology/Histotechnology Laboratory, Science
Applications International Corp., National Cancer
Institute at Frederick, Maryland 21702-1201, USA
||Department of Dermatopathology, Armed Forces
Institute of Pathology, Washington DC 20306, USA

1. Holman, C. D. J., Armstrong, B. K. & Heenan, P. J. *J. Natl Cancer Inst.* **71**, 651–656 (1983).
2. International Agency for Research on Cancer *IARC Monogr.* **55** (1992).
3. Whiteman, D. C., Whiteman, C. A. & Green, A. C. *Cancer Causes Cont.* **12**, 69–82 (2001).
4. Tietze, M. K. & Chin, L. *Mol. Med. Today* **6**, 408–410 (2000).
5. Takayama, H. et al. *Proc. Natl Acad. Sci. USA* **94**, 701–706 (1997).
6. Otsuka, T. et al. *Cancer Res.* **58**, 5157–5167 (1998).
7. Natali, P. G. et al. *Br. J. Cancer* **68**, 746–750 (1993).
8. Jeffers, M., Rong, S. & Vande Woude, G. F. *J. Mol. Med.* **74**, 505–513 (1996).
9. Chin, L., Merlino, G. & DePinho, R. A. *Genes Dev.* **12**, 3467–3481 (1998).
10. Noonan, F. P., Otsuka, T., Bang, S., Anver, M. R. & Merlino, G. *Cancer Res.* **60**, 3738–3743 (2000).
11. Hirobe, T. *Development* **102**, 567–574 (1988).
12. Noonan, F. P. & De Fabo, E. C. *Photochem. Photobiol.* **52**, 801–810 (1990).
13. De Fabo, E. C., Noonan, F. P. & Frederick, J. E. *Photochem. Photobiol.* **52**, 811–817 (1990).

Supplementary information is available at <http://www.nature.com> or as paper copy from the London editorial office of *Nature*.

stability of differentiation was confirmed by analysing several samples collected within sea areas (North Sea and Baltic Sea) in different years and seasons (see supplementary information).

We assigned individual cod to populations using a statistical procedure⁵ that attempts to minimize disequilibrium (Hardy–Weinberg and linkage) within groupings. This method can incorporate different assumptions about the level of migration between populations. We first assigned baseline samples using a parameter range that is likely to include the true migration rates, which are typically 1–10% (Table 1). The percentage of individuals with the highest probability of being derived from their population of origin was 97–100%, depending on population and on prior assumptions of migration (Table 1). Two ‘unknown’ samples (not included in the baseline data), each consisting of 30 individuals from the North Sea and the Baltic Sea, respectively, were assigned with similarly high precision to the samples from their populations of origin (Table 1).

Consequently, the reliability of assignment depends little on prior assumptions of migration rates or on whether the assigned individuals are included in the baseline data, a fact that demonstrates the robustness of the data. In fact, assignment is so reliable that testing as few as two or three individuals can provide an unambiguous conclusion about the origin of a sample that is claimed to originate from any one of the three sampled populations, representing the majority of cod landings in the northeastern Atlantic. Our method can readily be extended to include other cod populations (for example, from the northwestern Atlantic) by adding more baseline samples. Similar analyses should prove valuable in ensuring that exploitation of fish stocks is sustainable and that genetic resources are conserved in commercially important marine species.

Einar E. Nielsen*, **Michael M. Hansen***,
Cathrin Schmidt*†, **Dorte Meldrup***,
Peter Grønkjær‡

*Department of Inland Fisheries, Danish Institute
for Fisheries Research, 8600 Silkeborg, Denmark
e-mail: een@dfu.min.dk

†Present address: Institute for Hydrobiology and
Fisheries Science, 22767 Hamburg, Germany

‡Department of Marine Ecology, University of
Aarhus, DK-8200 Aarhus, Denmark

Fisheries

Population of origin of Atlantic cod

Most of the world’s cod (*Gadus morhua*) fisheries are now tightly regulated or closed altogether. Being able to link individual fish to their population of origin would assist enormously in policing regulations and in identifying poachers. Here we show that microsatellite genetic markers can be used to assign individual cod from three different populations in the northeastern Atlantic Ocean to their population of origin.

Genetic methods have been widely used for species identification of threatened marine fish¹ and mammals². Until now, however, no attempts have been made to link individual marine fish to their population of origin, presumably because of the low level of genetic differentiation among most marine fish populations³, and no

methods were available to determine the population of origin of individual cod in commercial landings or at fish markets.

We studied the three principal cod populations found in the northeastern Atlantic — from the North Sea, the Baltic Sea and the northeastern Arctic Ocean. As these populations are below safe biological thresholds⁴, fishing pressure is strictly regulated on an international level. For North Sea and Baltic Sea cod populations in particular, the proximity of populations and the number of nations that share quotas make poaching extremely difficult to control.

We collected tissue samples between 1995 and 1999. Genetic analysis of nine microsatellite markers revealed strong genetic differentiation between populations (F_{st} values (95% confidence interval): North Sea/Baltic Sea: 0.045 (0.042–0.056); North Sea/northeast Arctic Ocean: 0.035 (0.033–0.049); Baltic Sea/northeast Arctic Ocean: 0.073 (0.067–0.089)). The temporal

Table 1 Number of cod assigned to their population of origin

Migration rate (%)	North Sea	Baltic Sea	Northeastern Arctic Ocean
1	156 (99%)	152 (99%)	69 (100%)
5	156 (99%)	152 (99%)	69 (100%)
10	153 (97%)	151 (98%)	68 (99%)
N	158	154	69
Unknown samples	28 (93%)	30 (100%)	n/a

Values are numbers (percentages in parentheses) of individuals with the highest probability of belonging to the population from which they were sampled, assuming three different migration rates (baseline samples). N, number of individuals sampled from each population. ‘Unknown samples’ consist of 30 individuals from the North Sea or the Baltic Sea and were not included in the baseline samples. n/a, not available.

1. Ward, R. D. *Hydrobiologia* **420**, 191–201 (2000).
2. Baker, C. S. & Palumbi, S. R. *Science* **265**, 1538–1539 (1994).
3. Ward, R. D., Woodward, M. & Skibinski, D. O. F. *J. Fish Biol.* **44**, 213–232 (1994).
4. International Council for the Exploration of the Sea (ICES) *ACFM Report 2000* (<http://www.ices.dk/committe/acfm/comwork/report/2000/contents.html>).
5. Pritchard, J. K., Stephens, M. & Donnelly, P. *Genetics* **155**, 945–959 (2000).

Supplementary information is available at <http://www.nature.com> or as paper copy from the London editorial office of *Nature*.

Controlling photons using electromagnetically induced transparency

M. D. Lukin* & A. Imamoglu†‡

* Physics Department and ITAMP, Harvard University, Cambridge, Massachusetts 02138, USA

† Department of Electrical and Computer Engineering, University of California, Santa Barbara, California 93106, USA

‡ Faculty of Engineering and Natural Sciences, Sabanci University, Orhanli, 81474 Tuzla, Istanbul, Turkey

It is well known that a dielectric medium can be used to manipulate properties of light pulses. However, optical absorption limits the extent of possible control: this is especially important for weak light pulses. Absorption in an opaque medium can be eliminated via quantum mechanical interference, an effect known as electromagnetically induced transparency. Theoretical and experimental work has demonstrated that this phenomenon can be used to slow down light pulses dramatically, or even bring them to a complete halt. Interactions between photons in such an atomic medium can be many orders of magnitude stronger than in conventional optical materials.

Photons normally behave as non-interacting particles. This property ensures that information encoded in optical signals will be insensitive to environmental disturbances. As a consequence, optics has emerged as the preferred method for communicating information. In contrast, the processing of information requires interactions between signal carriers, that is, either between different photons or photons and electrons. Many other applications of optics, from medicine to spectroscopy, also rely on strong light–matter interactions. One of the main challenges of nonlinear optical science is the ‘tailoring’ of material properties to enhance such interactions, while minimizing the role of destructive processes such as photon absorption.

Electromagnetically induced transparency

The strength of the interaction between light and atoms is a function of the wavelength or frequency of light. When the light frequency matches the frequency of a particular atomic transition, a resonance condition occurs and the optical response of the medium is greatly enhanced. Light propagation is then accompanied by strong absorption and dispersion², as the atoms are actively promoted into fluorescing excited states.

Consider next the situation in which the atoms have a pair of lower energy states ($|1\rangle$ and $|2\rangle$ in Fig. 1a) in each of which the atoms can live for a long time. Such is the case for sublevels of different angular momentum (spin) within the electronic ground state of alkali atoms. In order to modify the propagation through this atomic medium of a light field that couples the ground state $|1\rangle$ to an electronically excited state $|3\rangle$ (red arrow), one can apply a second ‘control’ field that is near resonance with the transition $|3\rangle \rightarrow |2\rangle$ (black arrow). The combined effect of these two fields is to stimulate the atoms into a so-called coherent superposition of the states $|1\rangle$ and $|2\rangle$. Here the atoms can simultaneously occupy both states $|1\rangle$ and $|2\rangle$ with a definite phase relation between the two. In such a case, the two possible pathways in which light can be absorbed by atoms ($|1\rangle \rightarrow |3\rangle$ and $|2\rangle \rightarrow |3\rangle$) can interfere and cancel each other. With such destructive quantum interference, none of the atoms are promoted to the excited states, leading to a vanishing light absorption^{3–5}. This is the essence of electromagnetically induced transparency (EIT) and so-called ‘dark resonances’ see refs 2–4).

Propagation in EIT medium

Many of the important properties of EIT result from the fragile

nature of quantum interference in a material that is initially opaque. Indeed, the ideal transparency is attained only if the frequency difference between the two laser fields precisely matches the frequency separation between the two lower states. If matching is not perfect, the interference is not ideal and the medium becomes absorbing. Hence the transparency spike that appears in the absorption spectrum is typically very narrow (Fig. 1b). The tolerance to frequency mismatch can be increased by using stronger coupling fields, because then interference becomes more robust.

Atoms are decoupled from the light fields in an ideal EIT, so the refractive index at resonance is nearly equal to unity. This means that the propagation velocity of a phase front (that is, the phase velocity) is equal to that in vacuum. However, the narrow transparency resonance is accompanied by a very steep variation of the refractive index with frequency. As a result, the envelope of a wavepacket propagating in the medium moves with a group velocity v_g (ref. 6) that is much smaller than the speed of light in vacuum, c . We note that v_g depends on the control field intensity and the atomic density: decreasing the control power or increasing the atom density makes v_g slower.

Figure 2 illustrates the dynamics of light propagation in EIT medium. Initially the pulse is outside the medium in which all atoms are in their ground states ($|1\rangle$). The front edge of the pulse then enters the medium and is rapidly decelerated. Because it is still outside the medium, the back edge propagates with vacuum speed c . Thus, upon entrance into the cell the spatial extent of the pulse is

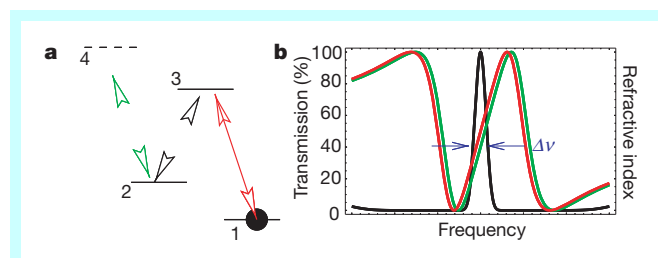


Figure 1 Electromagnetically induced transparency. **a**, Prototype atomic system for EIT and nonlinear optics. **b**, Spectrum of transmission and refractive index corresponding to EIT. Rapid variation of the refractive index (red curve) causes a reduction of group velocity. A control field (black arrow) is used to modify the propagation of weak resonant field (red) or to induce its interaction with another weak field (green). The presence of a second weak field causes an effective shift of the resonant frequency (green curve), which results in a corresponding change of the refractive index.

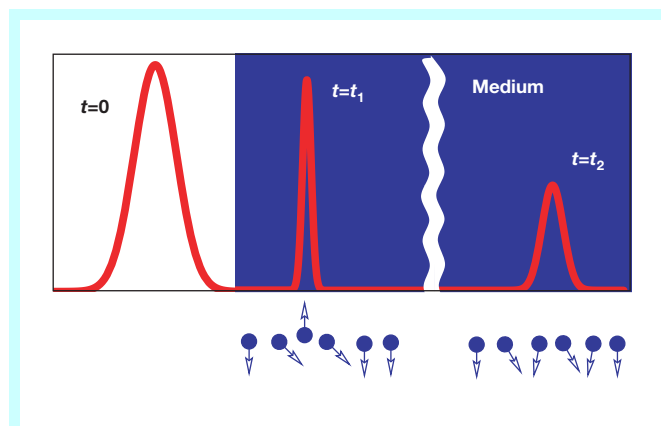


Figure 2 Schematic of spatial compression exhibited when a light pulse (red curve) enters the slow medium (blue). Photons are converted into flipped spins (blue arrowed circles), and the slow photonic and spin waves then propagate together. For long distances ($t_2 \gg t_1$), the lossless propagation is limited by the spreading of the pulses owing to the narrow bandwidth of the transparency window.

compressed by the ratio c/v_g , whereas its peak amplitude remains unchanged. Clearly the energy of the light pulse is much smaller when it is inside the medium. The rest of the photons are being expended to establish the coherence between the states $|1\rangle$ and $|2\rangle$, or in other words to flip atomic spins, with any excess energy carried away by the control field. The wave of flipped spins now propagates together with the light pulse. The two form a combined excitation of photons and spins called a dark-state polariton⁷. The group velocity of this polariton is proportional to the magnitude of its photonic component. As the pulse exits the medium its spatial extent increases again and the atoms return to their original ground state; the pulse however, is delayed as a whole by $\tau = (1/v_g - 1/c)L$, where L is the length of the medium.

The above description is an ideal scenario. In practice, two main limitations have to be taken into account. First, the lifetime of the coherence created between different spin states is always finite (τ_{coh}). This implies that after a characteristic time τ_{coh} the atoms will end up either in state $|1\rangle$ or $|2\rangle$; the transparency will then be lost and the polariton will disappear. Second, even for infinite τ_{coh} , the pulse delay is limited by the bandwidth of the transparency window, which decreases with propagation distance. This is because at higher densities or propagation distances the medium becomes increasingly opaque at frequencies other than the line centre; as a result the available transparency window becomes smaller. After a sufficiently long propagation this results in spreading of the pulse (see Fig. 2). In order to preserve the pulse, its bandwidth $1/T$ should be smaller than the transparency bandwidth $\Delta\nu$. This puts a limit on the ratio of delay and pulse duration, which can exceed unity only if an optically dense medium is used (see Box 1).

Early experimental work demonstrated these unusual properties of EIT^{8–11}. For example, group velocities of $c/165$ were measured in lead vapour⁹. Even in these early experiments, delays considerably exceeding the pulse duration were observed. The subject was then brought into focus by the remarkable experiments of ref. 12. An ultracold gas of Na atoms was used to slow light pulses to 17 m per second. The pulses were entirely localized inside the medium owing to its large density. Experimental work on slow group velocities in hot atomic vapour^{13,14} rapidly followed.

Nonlinear optics based on EIT

The large body of work on EIT was motivated by the realization that nonlinear optical effects can be enhanced when resonant absorption is eliminated. The key idea¹⁵ is that in an EIT medium the atoms are very sensitive to processes that create or alter the coherent

Box 1

Slow velocities versus narrow lines

Consider a generic dispersive optical element (for example, narrow atomic resonance, optical cavity, and so on) that is characterized by the common lorentzian input–output relation: $E_{\text{out}}(\omega) = T_1(\omega)E_{\text{in}}(\omega)$ with spectral response function $T_1 = (1 + i(\omega - \omega_{\text{res}})/\Delta\nu)^{-1}$. Here, $\Delta\nu$ is the linewidth of the dispersive element and ω_{res} is the resonance frequency of the optical element. For $\omega \approx \omega_{\text{res}}$, this response corresponds to a group delay $\tau_1 = 1/\Delta\nu$. Hence narrow resonances lead to long group delays, but the latter does not exceed the reciprocal linewidth. In contrast, the response near an EIT resonance is given by¹¹: $T_{\text{EIT}}(\omega) = \exp(i\omega\tau - \omega^2/\Delta\nu^2 + O(\omega^3))$, with the corresponding bandwidth $\Delta\nu = 1/\tau \times \sqrt{T}$. Here $T = N\sigma L/2$ is the optical depth of the medium, with N and σ being the atomic density and absorption cross-section.

The importance of a product with a large-delay bandwidth can be emphasized by noting that the spatial length of light pulse inside a medium is $v_g T$. Hence in order to localize the pulse inside a medium of length L and to contain its spectrum within the transparency linewidth, $\Delta\nu \times \tau \gg 1$ is required.

superposition between states $|2\rangle$ and $|1\rangle$. As an illustration, we consider the process where an optical field with amplitude E_s (green arrow in Fig. 1a) is used to modify the properties of the medium by coupling to a third optically allowed transition between states $|2\rangle$ and $|4\rangle$. This off-resonant field effectively changes the energy of the metastable state $|2\rangle$ by an amount that is proportional to the intensity E_s^2 . This shift, in turn, modifies the refractive index of the probe field. When the dispersion of refractive index is very steep (Fig. 1b), small energy shifts due to weak optical fields result in a large index change. This is the essence of the resonantly enhanced Kerr effect¹⁶.

Modification of the refractive index of one field by another can be viewed as resulting from a photon–photon interaction that preserves the intensity of each field mode. This process is also referred to as cross-phase modulation: the two-field interaction gives rise to a nonlinear phase shift that is proportional to the group delay (that is, the interaction time of the slow pulse with matter). Hence for given field intensities, increasing the interaction time between light and matter can result in more efficient nonlinear optical effects.

This is just an example of how nonlinear optics based on EIT works. By now related ideas have been explored for a variety of nonlinear phenomena ranging from phase conjugation¹⁷ to acousto-optics¹⁸ and have already been shown to work very well for a number of potential practical applications. This is perhaps best illustrated by experimental work¹⁹ that used EIT to obtain extremely efficient frequency up-conversion into the extreme ultraviolet. An extension of this work to molecular systems allowed the generation of a comb of phase modulated signals with a frequency span corresponding to subfemtosecond pulses^{20,21}.

We note in particular the recent work on nonlinear optics involving very weak light fields. For example, in the experiment on slow light propagation¹², an optical Kerr effect that is several orders of magnitude stronger than that in usual resonant systems was inferred. In another recent experiment²², an efficient nonlinear process involving mixing of four waves was initiated by optical pulses with tiny energies in the nanojoule range.

Given the rapid progress of studies on EIT-based nonlinear optics it is natural to ask about the ultimate limits. For example, in cross-phase modulation¹⁶ how much energy in one pulse is required, under optimal conditions, to change the phase of another pulse by π ? This problem was studied²³ for the case involving one ‘slow’ pulse that is subject to EIT and another ‘fast’ pulse that propagates at c . In this case the interaction time is limited by the duration of the fast pulse (T), and the efficiency of various processes scales proportionally to the

number of photons n in the pulse per cross-section of a single atom. Specifically, an optical pulse containing few photons can be used to change significantly the phase of another field. Ref. 24 extended these studies by considering two interacting slow pulses. In this case, the interaction time is determined solely by the group delay τ . As τ can exceed the pulse length T in an optically dense medium (see Box 1), this approach in principle allows one to induce a large nonlinear phase shift, even when the energy of each input pulse is less than that of a single photon per atomic cross-section.

From EIT to quantum memory

It is now widely accepted that quantum mechanics allows for fundamentally new forms of communication and computation²⁵. To implement these ideas, information should be encoded in delicate quantum states, such as single-photon states, and subsequently manipulated without being destroyed. Photons are the fastest and simplest carriers of quantum information, but they are difficult to localize and process. It appears that an ideal solution would be to store and process quantum information in matter that forms the nodes of a quantum network, and to communicate between these nodes using photons²⁶.

A way of achieving this goal by transferring or mapping quantum information reversibly between light and matter was proposed in ref. 27. The idea is closely related to the dark-state polariton concept⁷. When a polariton propagates in an EIT medium, its properties can be modified simply by changing the intensity of the control beam. As the control intensity is decreased the group velocity is slowed, which also implies that the contribution of photons in the polariton becomes purely atomic, and its group velocity is reduced to zero. At this point, quantum information originally carried by photons is mapped onto long-lived spin states of atoms. As long as the trapping process is sufficiently smooth (that

is, adiabatic²⁸), the entire procedure has no loss and is completely coherent. The stored quantum state can easily be retrieved by simply re-accelerating the stopped polariton. This is illustrated in Fig. 3, which shows the evolution of the 'signal' light pulse, spin coherence and polariton when the control beam is turned off and on. The amplitude of the signal pulse decreases as it is being decelerated whereas the spin coherence grows; the procedure is reversed when the control beam is turned back on.

We note here that the essential point of this technique is not to store the energy or momentum carried by photons but to store their quantum states. In fact, in practice almost no energy or momentum is actually stored in the EIT medium. Instead, both are being transferred into (or borrowed from) the control beam in such a way that an entire optical pulse is coherently converted into a low-energy spin wave. This is the key feature that distinguishes the EIT approach from earlier studies in optics (involving, for example, traditional photon echo techniques¹) or nuclear physics²⁹; it also makes possible applications in quantum information science. A different technique, to 'freeze' light pulses, was suggested in ref. 30.

We have already argued that for EIT to be effective in eliminating dissipation, the light pulse spectrum should be contained within a relatively narrow transparency window (Fig. 1b). A vanishing control beam intensity implies that the transparency window would become infinitely narrow and eventually disappear. The essence of adiabatic following in polaritons is that a dynamic reduction in group velocity is accompanied by narrowing of the polariton frequency spectrum, such that it is not destroyed even if $v_g = 0$. The conditions for adiabatic following are very simple: the entire pulse should be within the medium at the beginning of the trapping procedure and its spectrum should be contained within the original transparency window. Once again, these conditions are satisfied only if the medium is optically dense.

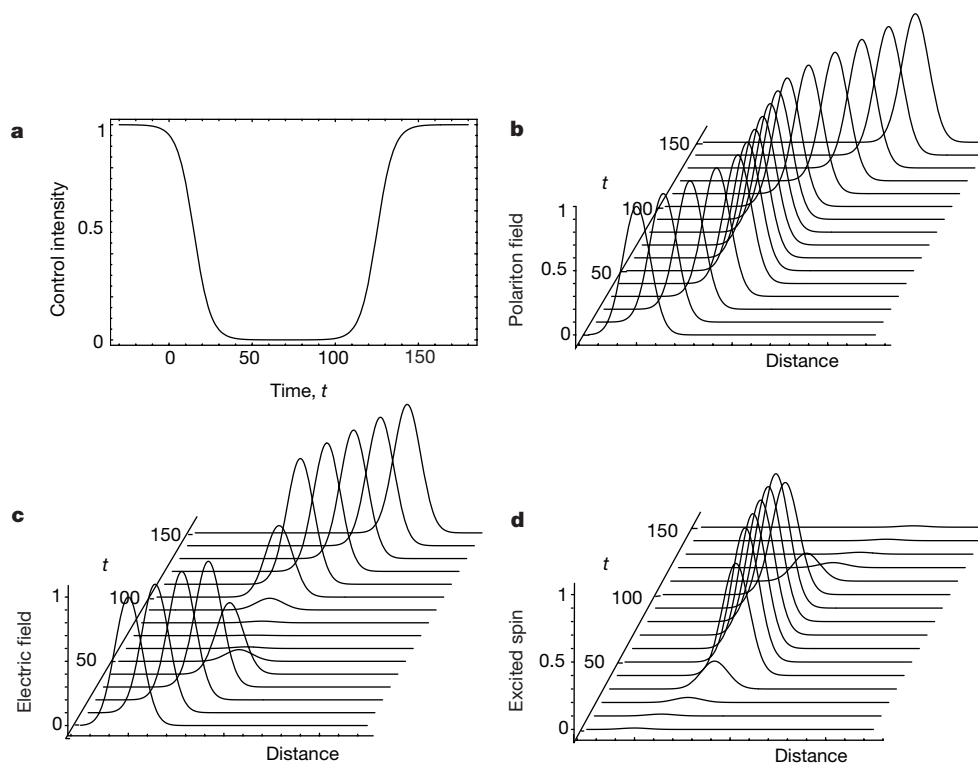


Figure 3 Dark-state polaritons. A dark-state polariton can be stopped and re-accelerated by ramping the control field intensity as shown in **a**. The coherent amplitudes of the polariton Ψ , the electric field E and the spin components S are plotted in **b** to **d**.

Recent experiments have already verified the basic concept of dynamic trapping in polaritons, by showing that the weak laser pulses are not destroyed during the trapping and release processes. In the experiment of ref. 31, a light pulse was slowed and then 'trapped' in an ultracold atomic sample for up to 1.5 ms. The experiment of ref. 32 used Zeeman sublevels of the electronic ground state of warm Rb atoms to demonstrate a trapping time of about 0.5 ms. Extension of the latter work recently demonstrated the phase coherence of the trapping process³³.

Outlook

Even though EIT has already made a major impact in nonlinear optical science, commercial applications have not yet emerged. One potential area is all-optical switching and signal processing in optical communication. The most serious roadblocks on this front are materials and speed issues. Good optical control requires long coherence times, and for this reason the majority of experiments made use of atomic vapours that have relatively slow response. For practical communication systems, solid state devices are desirable because of their low cost and the possibility of integration with existing technologies. Work on implementing EIT in solids is currently in progress^{34,35}.

Photon-photon interactions enabled by EIT can fulfil the stringent requirements on precision and efficiency imposed by quantum information processing. In particular, optical materials with large nonlinearities and low loss could be indispensable for the controlled generation of entangled states and for quantum logic operations. Several avenues for using EIT in this area have already been explored. Earlier proposals involved the use of a coherent medium to enhance photon-photon interactions in optical cavities. The key idea is that a single photon can shift the resonant frequency such that the following photon is out of resonance and is therefore reflected³⁶. The resulting 'photon blockade' effect can form the basis of a quantum switch. However, the requirement of a high-quality cavity is a disadvantage from a practical point of view. Subsequent work has predicted the efficient generation of entangled photons on the basis of resonant mixing of four waves³⁷. Using EIT-based phase modulation for two slowly propagating pulses, ref. 24 predicted the possibility of generating macroscopic quantum states (so-called 'Schrödinger's cat' states) of light. However, the application of this idea to quantum logic operations is complicated by the evolution of pulse envelopes in nonlinear process. Nevertheless, a scheme for complete quantum teleportation using this technique has recently been proposed³⁸.

We also note that once a dark-state polariton is converted into a purely atomic excitation in a small-sized sample, logic operations can be accomplished by promoting atoms into excited states with strong atom-atom interactions, as suggested in ref. 39. Here the ability to interconvert quantum information between photons and atoms is essential for performing operations involving distant units and for the scalability of such systems. Likewise, important potential applications such as quantum communications over long distances might be implemented by combining an EIT-based memory with linear optical elements⁴⁰.

Experiments in the coming years will undoubtedly shed new light on the opportunities and practical limitations for classical signal processing and for quantum-state manipulation using EIT. Nevertheless, it seems safe to predict that some of the techniques described here will find important uses in optical science and technology. □

1. Boyd, R. W. *Nonlinear Optics* (Academic, Boston, 1992).

2. Scully, M. O. & Zubairy, M. S. *Quantum Optics* (Cambridge Univ. Press, Cambridge, 1997).

3. Arimondo, E. in *Progress in Optics* (ed. Wolf, E.) Vol. 35, 259–354 (North-Holland, Amsterdam, 1996).

4. Harris, S. E. Electromagnetically induced transparency. *Phys. Today* 50(7), 36–42 (1997).

5. Boller, K., Imamoglu, A. & Harris, S. E. Observation of an electromagnetically induced transparency. *Phys. Rev. Lett.* 66, 2593–2596 (1991).
6. Harris, S. E., Field, J. E. & Kasapi, A. Dispersive properties of electromagnetically induced transparency. *Phys. Rev. A* 46, R29–R32 (1992).
7. Fleischhauer, M. & Lukin, M. D. Dark-state polaritons in electromagnetically induced transparency. *Phys. Rev. Lett.* 84, 5094–5097 (2000).
8. Xiao, M., Li, Y., Jin, S. & Gea-Banacloche, J. Measurement of dispersive properties of electromagnetically induced transparency in rubidium atoms. *Phys. Rev. Lett.* 74, 666–669 (1995).
9. Kasapi, A., Jain, M., Yin, G. Y. & Harris, S. E. Electromagnetically induced transparency: Propagation dynamics. *Phys. Rev. Lett.* 74, 2447–2450 (1995).
10. Schmidt, O., Wynands, R., Hussein, Z. & Meschede, D. Steep dispersion and group velocity below $c/3000$ in coherent population trapping. *Phys. Rev. A* 53, R27–R30 (1996).
11. Lukin, M. D. *et al.* Spectroscopy in dense coherent media: line narrowing and interference effects. *Phys. Rev. Lett.* 79, 2959–2962 (1997).
12. Hau, L. V., Harris, S. E., Dutton, Z. & Behroozi, C. H. Light speed reduction to 17 metres per second in an ultracold atomic gas. *Nature* 397, 594–598 (1999).
13. Kash, M. M. *et al.* Ultra-slow group velocity and enhanced nonlinear optical effects in a coherently driven hot atomic gas. *Phys. Rev. Lett.* 82, 5229–5232 (1999).
14. Budker, D. *et al.* Nonlinear magneto-optics and reduced group velocity of light in atomic vapor with slow ground state relaxation. *Phys. Rev. Lett.* 83, 1767–1770 (1999).
15. Harris, S. E., Field, J. E. & Imamoglu, A. Nonlinear optical processes using electromagnetically induced transparency. *Phys. Rev. Lett.* 64, 1107–1110 (1990).
16. Schmidt, H. & Imamoglu, A. Giant Kerr nonlinearities obtained by electromagnetically induced transparency. *Opt. Lett.* 21, 1936–1938 (1996).
17. Hemmer, P. *et al.* Efficient low-intensity optical phase conjugation based on coherent population trapping in sodium. *Opt. Lett.* 20, 982–984 (1995).
18. Matsko, A. B. *et al.* Anomalous stimulated Brillouin scattering via ultraslow light. *Phys. Rev. Lett.* 86, 2006–2009 (2001).
19. Jain, M. *et al.* Efficient nonlinear frequency conversion with maximal atomic coherence. *Phys. Rev. Lett.* 77, 4326–4329 (1996).
20. Sokolov, A. V. *et al.* Raman generation by phased and antiphased molecular states. *Phys. Rev. Lett.* 85, 562–565 (2000).
21. Liang, J. Q. *et al.* Sideband generation using strongly driven Raman coherence in solid hydrogen. *Phys. Rev. Lett.* 85, 2474–2477 (2000).
22. Zibrov, A. S., Lukin, M. D. & Scully, M. O. Non-degenerate parametric self-oscillation via resonantly enhanced multiwave mixing in coherent media. *Phys. Rev. Lett.* 83, 4049–4052 (1999).
23. Harris, S. E. & Hau, L. V. Nonlinear optics at low light levels. *Phys. Rev. Lett.* 82, 4611–4614 (1999).
24. Lukin, M. D. & Imamoglu, A. Nonlinear optics and quantum entanglement of ultra-slow single photons. *Phys. Rev. Lett.* 84, 1419–1422 (2000).
25. Bouwmeester, D., Ekert, A. & Zeilinger, A. *The Physics of Quantum Information* (Springer, Berlin, 2000).
26. Cirac, J. L., Zoller, P., Kimble, H. J. & Mabuchi, H. Quantum state transfer and entanglement distribution among distant nodes in a quantum network. *Phys. Rev. Lett.* 78, 3221–3224 (1997).
27. Lukin, M. D., Yelin, S. F. & Fleischhauer, M. Entanglement of atomic ensembles by trapping correlated photon states. *Phys. Rev. Lett.* 84, 4232–4235 (2000).
28. Oreg, J., Hioe, F. T. & Eberly, J. H. Adiabatic following in multilevel systems. *Phys. Rev. A* 29, 690–697 (1984).
29. Shvyd'ko, Yu. V. *et al.* Storage of nuclear energy through magnetic switching. *Phys. Rev. Lett.* 77, 3232–3235 (1996).
30. Kocharovskaya, O., Rostovtsev, Yu. & Scully, M. Stopping light via hot atoms. *Phys. Rev. Lett.* 86, 628–631 (2001).
31. Liu, C., Dutton, Z., Behroozi, C. H. & Hau, L. V. Observation of coherent optical information storage in an atomic medium using halted light pulses. *Nature* 409, 490–493 (2001).
32. Phillips, D., Fleischhauer, A., Mair, A., Walsworth, R. & Lukin, M. D. Storage of light in atomic vapor. *Phys. Rev. Lett.* 86, 783–786 (2001).
33. Mair, A. *et al.* Phase coherence and control of stored photonic information. *Phys. Rev. Lett.* (submitted).
34. Ham, B., Shahriar, M. S. & Hemmer, P. Enhanced nondegenerate four-wave mixing owing to electromagnetically induced transparency in a spectral hole-burning crystal. *Opt. Lett.* 22, 1138–1140 (1997).
35. Imamoglu, A. Electromagnetically induced transparency with two dimensional electron spins. *Opt. Comm.* 179, 179–182 (2000).
36. Imamoglu, A., Schmidt, H., Woods, G. & Deutsch, M. Strongly interacting photons in a nonlinear cavity. *Phys. Rev. Lett.* 79, 1467–1470 (1997).
37. Lukin, M. D., Matsko, A., Fleischhauer, M. & Scully, M. O. Quantum noise and correlations in resonantly enhanced wave mixing based on atomic coherence. *Phys. Rev. Lett.* 82, 1847–1850 (1999).
38. Vitali, D., Fortunato, M. & Tombesi, P. Complete quantum teleportation with a Kerr nonlinearity. *Phys. Rev. Lett.* 85, 445–448 (2000).
39. Lukin, M. D. *et al.* Dipole blockade and quantum information processing in mesoscopic atomic ensembles. *Phys. Rev. Lett.* 87, 037901 (2001).
40. Duan, L. M., Lukin, M. D., Cirac, J. I. & Zoller, P. Long-distance quantum communication with atomic ensembles and linear optics. Preprint quant-ph/0105105 at <http://xxx.lanl.gov> (2001); *Nature* (submitted).

Acknowledgements

We would like to thank S. Harris and M. Scully for many discussions. We also thank J. I. Cirac, L.-M. Duan, M. Fleischhauer, D. Phillips, H. Schmidt, R. Walsworth, S. Yelin and P. Zoller. This work was supported by the NSF via a grant to ITAMP and through an ITR program (M.L.) and by a David and Lucile Packard Fellowship (A.I.).

Correspondence and requests for materials should be addressed to A.I. (e-mail: atac@ece.ucsb.edu).

Skeletons of terrestrial cetaceans and the relationship of whales to artiodactyls

J. G. M. Thewissen*, E. M. Williams*, L. J. Roe† & S. T. Hussain‡

* Department of Anatomy, Northeastern Ohio Universities College of Medicine, Rootstown, Ohio 44272, USA

† Department of Geosciences, University of Arizona, Tucson, Arizona 85721, USA

‡ Department of Anatomy, Howard University, College of Medicine, Washington DC 20059, USA

Modern members of the mammalian order Cetacea (whales, dolphins and porpoises) are obligate aquatic swimmers that are highly distinctive in morphology, lacking hair and hind limbs, and having flippers, flukes, and a streamlined body. Eocene fossils document much of cetaceans' land-to-water transition, but, until now, the most primitive representative for which a skeleton was known was clearly amphibious and lived in coastal environments. Here we report on the skeletons of two early Eocene pakicetid cetaceans, the fox-sized *Ichthyolestes pinfoldi*, and the wolf-sized *Pakicetus attocki*. Their skeletons also elucidate the relationships of cetaceans to other mammals. Morphological cladistic analyses have shown cetaceans to be most closely related to one or more mesonychians, a group of extinct, archaic ungulates, but molecular analyses have indicated that they are the sister group to hippopotamids. Our cladistic analysis indicates that cetaceans are more closely related to artiodactyls than to any mesonychian. Cetaceans are not the sister group to (any) mesonychians, nor to hippopotamids. Our analysis stops short of identifying any particular artiodactyl family as the cetacean sister group and supports monophyly of artiodactyls.

In contrast to the debate about the cetacean sister group, the relationships among Eocene cetaceans and the content of Cetacea itself are not controversial^{1–5}. All phylogenetic studies indicate that pakicetids are more closely related to living cetaceans than to artiodactyls and mesonychians, and that pakicetids share the cetacean synapomorphies of the ear^{2,3,6}. Pakicetids are followed by ambuloctetids in the cladogram, and modern cetaceans (toothed and baleen whales) are closely related to late Eocene basilosaurids and dorudontids^{1,3–5}. The most archaic cetacean for which the skeleton is known is the amphibious *Ambulocetus natans*^{7,8}. It was a powerful, walrus-sized animal that lived in coastal environments and resembled a crocodile, with the exception of long hind limbs that were used in swimming⁹. Although *Ambulocetus* is unlike modern cetaceans, it also differs strongly from its land mammal relatives, be they artiodactyls or mesonychians. Adaptations for life in water in *Ambulocetus* and later whales complicate determination of their closest relatives among the land mammals. Data from fossil whales that are more basal on the cetacean phylogenetic tree and have fewer aquatic adaptations could presumably yield new phylogenetic insights^{10,11}. Pakicetids are in this position and can be used to test mesonychian and hippopotamid hypotheses.

There are three genera of pakicetid cetaceans: *Pakicetus*, *Nalacetus* and *Ichthyolestes*². *Pakicetus* is the largest, followed by *Nalacetus* (approximately 5% smaller in linear dimensions), and *Ichthyolestes* (approximately 29% smaller). Until now, only teeth, jaws and one braincase have been described for pakicetids^{2,3,12}. We excavated four partial skulls, two of which retain the orbital region, several snout fragments, and approximately 150 isolated postcranial bones of pakicetids from multiple individuals. These were found at a single site in the early Eocene Kuldana Formation of Pakistan. We use these fossils to show (1) that these archaic cetaceans were land mammals; and (2) that cetaceans are more closely related to artiodactyls than to mesonychians.

Pakicetid form and function

Aquatic postcranial adaptations are pronounced in late Eocene basilosaurids and dorudontids, the oldest obligate aquatic cetaceans for which the entire skeleton is known^{13–15}, and therefore can be used to evaluate pakicetid morphology. Aquatic adaptations of basilosaurids and dorudontids include: presence of short neck

vertebrae; thoracic and lumbar vertebrae that are similar in length; unfused sacral vertebrae; lack of a sacro-iliac joint; presence of a short tail with a ball-vertebra (a vertebra at the base of the fluke, with convex articular surfaces); broad fan-shaped scapula with anterior acromion and small supraspinous fossa; an ulna with a large and transversely flat olecranon; a wrist and distal forearm flattened in the plane of the hand; and tiny hind limbs¹⁵.

Pakicetids display none of these features. Pakicetid neck vertebrae are longer than late Eocene whales (Fig. 1a and b), and the trunk vertebrae increase in size from anterior to posterior (Fig. 1c–f), as in land mammals. Lumbar and caudal vertebrae (Fig. 1e–i) are long compared to those of modern fluked cetaceans, but not as long as in extinct cetaceans that swam by undulating their entire spine (for example, the remingtonocetid *Kutchicetus*¹⁶). *Ambulocetus* and *Kutchicetus* have a muscular and flexible lumbar vertebral column, whereas motion in pakicetids is restricted as a result of their revolute zygapophyses (Fig. 1e), a feature in common with stiff-backed runners such as mesonychians¹⁷ and many extinct and modern artiodactyls¹⁸. The pakicetid sacrum consists of four solidly fused vertebrae and there is a strong sacro-iliac joint, as in land mammals and in amphibious whales such as *Ambulocetus*^{8,19} and *Kutchicetus*¹⁶ but unlike later cetaceans¹⁴.

The pakicetid scapula (Fig. 2) has a large supraspinous fossa with a small acromion, unlike any other cetaceans^{13,15}. The humerus is long and slender (Fig. 1j and k), and all but lacks a deltopectoral crest, as in running mammals. This crest is large and reaches distally in modern sirenians and pinnipeds²⁰ as well as in other Eocene cetaceans^{13,15}. Distally, the humerus has a wide, tightly articulating hinge joint for the radius and ulna allowing a great degree of flexion but few other motions. This is unlike other cetaceans, but is common in running mammals²¹. The forearm of pakicetids is not transversely flattened. The olecranon (Fig. 1l and m) makes up less than 12% of the length of the ulna in *Ichthyolestes*, whereas the olecranon is large in *Ambulocetus* (24%). In swimmers, such as basilosaurids¹³ and pinnipeds²⁰, the olecranon is antero-posteriorly and proximo-distally long and provides a strong lever for elbow extension and wrist flexion. Short olecrana occur in runners²¹.

The pakicetid innominate (Fig. 1n) is large and the ischium is longer than the ilium. The tibia is long in pakicetids and has a short tibial crest. Long tibiae are present in fast land mammals²¹ and also

in phocid seals^{20,22}. In phocids, the large ischium, short femur with asymmetrical condyles combined with the long tibia with long tibial crest allow the hamstrings to act as foot adductors (knee flexors with thigh in abducted position) while swimming²⁰. The relative lengths of these bones in pakicetids, their slender appearance, the short tibial crest, high patellar groove, and symmetrical knee make the phocid mode of locomotion unlikely for pakicetids, and their external morphology is more similar to that of running and jumping mammals²¹.

Running features are also found in the ankle where the proximal trochlea of the astragalus is constrained to a tight hinge joint (Fig. 1o–r). Like artiodactyls^{23,24}, pakicetids have a trochleated astragalar head rotating in the dorso-plantar plane. The sustentacular facet is a hinge that also rotates dorso-plantarly, and the ectal facet is small and laterally placed. The calcaneum (Fig. 1s and t) has

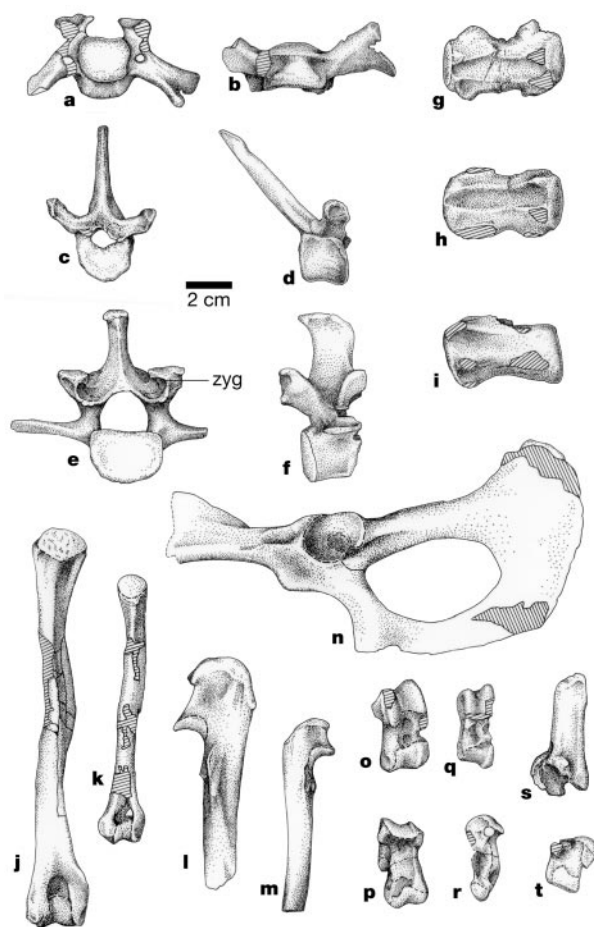


Figure 1 Postcranial osteology of pakicetids. H-GSP 96218, cervical vertebra of *Pakicetus* in anterior (a) and ventral (b) view. H-GSP 96516, thoracic vertebra of *Pakicetus* in anterior (c) and right lateral (d) view. H-GSP 98154, lumbar vertebra of *Pakicetus* in posterior (e) and right lateral (f) view. H-GSP 96564, caudal vertebra of *Pakicetus* in ventral view (g). H-GSP 96305, caudal vertebra of *Pakicetus* in ventral (h) and left lateral view (i). H-GSP 92042, left humerus of *Pakicetus*, subadult, in posterior view (j). H-GSP 30128, left humerus of *Ichthyolestes*, subadult, in posterior view (k). H-GSP 96057, proximal left ulna of *Pakicetus* in medial view (l). H-GSP 30286, proximal right ulna of *Ichthyolestes* in lateral view (m). H-GSP 98134, left innominate of *Pakicetus* in lateral view (n, complete ilium present in H-GSP 30288). H-GSP 98148, left astragalus of *Pakicetus* in dorsal (o) and plantar view (p). H-GSP 98149 left astragalus of *Ichthyolestes* in dorsal (q) and lateral (r) view. H-GSP 96359, right calcaneum of *Pakicetus* in medial view (s, distal part broken). H-GSP 96420, left distal calcaneum of *Pakicetus* in medial view (t). Abbreviation: zyg, zygapophysis of lumbar vertebra, illustrating recurving articular surface.

a long tuber and an obliquely set, narrow cuboid facet. These features are commonly interpreted as adaptations for running^{23–25}, although they are retained in non-running, graviportal artiodactyls.

The hands of *Pakicetus* and *Ambulocetus* are equally robust; the ratio of midshaft width to length of the central metacarpal is 0.17 and 0.18 respectively. On the other hand, the feet of *Ambulocetus* exceed those of *Pakicetus* in robustness by more than 20% (this ratio for metatarsal III is 0.11 and 0.14 respectively). *Ambulocetus* probably swam using its hind limbs as the main propulsor, and its robust feet may be an adaptation for forcefully displacing water during swimming⁹. Pakicetids, on the other hand, had the slender metapodials of running animals.

The cranial morphology of pakicetids is consistent with the evidence from the postcranium. The nasal opening of pakicetids was at the tip of the snout², as in land mammals and other primitive cetaceans²⁶, but unlike late Eocene cetaceans¹³ and other marine mammals²⁷ (sirenians, desmostylians). The lacrimal foramen is present in pakicetids (Fig. 3) and other archaic cetaceans, but is usually absent in aquatic mammals (modern cetaceans²⁷, sirenians²⁷ and pinnipeds²²).

The orbits of pakicetids are close together and are frontated (face dorsally) but are not at the most dorsal point of the head (Fig. 3). This is unlike any other cetacean. Most middle Eocene and all later cetaceans have orbits positioned below the supra-orbital shield and facing laterally^{13,28}, an adaptation for submerged living. The orbits of *Ambulocetus* are not frontated but are positioned dorsally⁸, as in modern amphibious mammals, such as hippopotami. The pakicetid position enhances binocular vision but is not necessarily related to life in water.

Deep, near-vertical gouges constitute most of the dental wear in pakicetids²⁹. Cladistic arguments have been used to link this wear pattern to aquatic predation on fish²⁹, but no functional model or modern analogue is known. Moreover, this kind of dental wear also occurs in raoellid artiodactyls³⁰. Although this dental wear probably represents a distinctive way of food processing, it does not necessarily imply aquatic life.

Unlike any other cetacean, the pakicetid outer ear was unspecialized and similar to that of land mammals⁶. The external auditory meatus opens low on the side of the skull, and the mandible has a small mandibular foramen³¹. In amphibious mammals, the external auditory meatus commonly opens dorsally. The mandibular foramen of late Eocene and Neogene cetaceans is large^{13,15} and transmits underwater sound to the middle ear. Enlargement also occurs in *Ambulocetus*⁸, but the foramen is small in pakicetids^{2,31}.

The pakicetid middle ear was highly specialized and included pachy-osteosclerotic ossicles², an involucrum⁶ and a plate-like sigmoid process⁶. These features have been interpreted as adaptations for underwater hearing³¹, and it has been suggested that the presence of an involucrum facilitates underwater high-frequency transmission in modern odontocetes³² even though the involucrum is also present in low-frequency mysticetes. In the case of pakicetids, the absence of air sinuses insulating the ears¹², the firm fusion of the periotic to the surrounding bones^{2,12}, and the presence of a flat tympanic membrane^{3,6} suggest that reception of airborne sound is well developed, but are inconsistent with good underwater hearing^{3,12}. It is most likely that the specializations of the pakicetid middle ear are analogous to those of some subterranean mammals³³ and are related to the reception of substrate-borne vibrations or sound when the head is in contact with the ground⁸. Turtles are in close contact with the substrate and gather sensory information using this method³⁴.

Taken together, the features of the skull indicate that pakicetids were terrestrial, and the locomotor skeleton displays running adaptations. Some features of the sense organs of pakicetids are also found in aquatic mammals, but they do not necessarily imply life in water. Pakicetids were terrestrial mammals, no more amphibious than a tapir.

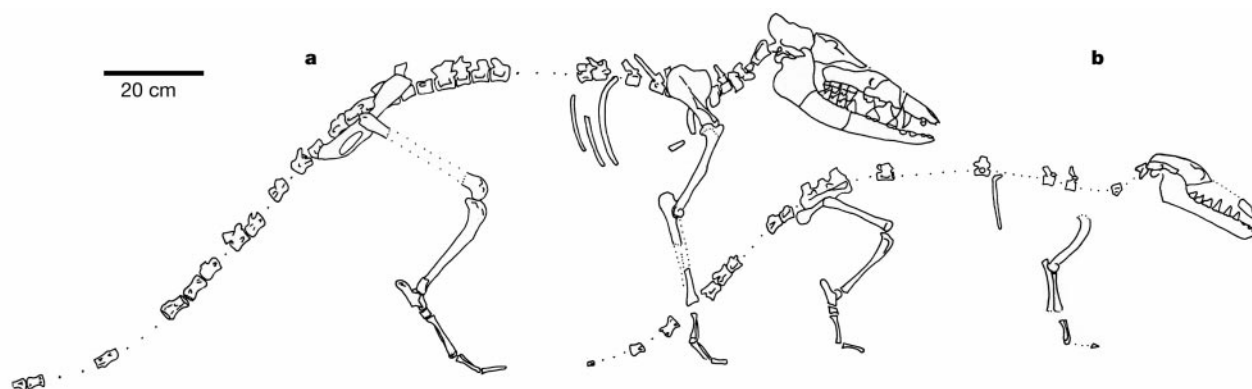


Figure 2 Skeletons of the pakicetid cetaceans *Pakicetus* (a) and *Ichthyolestes* (b). Reconstructions are based on fossils from H-GSP Locality 62 in the Eocene of Pakistan. Unknown elements have not been reconstructed.

Phylogenetic analysis

There are two current hypotheses about the closest relatives of cetaceans, championed by morphological and molecular systematists respectively. Morphological phylogenetic studies^{3,4,35–37} indicate that the sister group to cetaceans is (one of) the mesonychians, an extinct group of flesh-eating ungulates. On the other hand, molecular studies^{38–41} indicate that cetaceans are embedded in paraphyletic artiodactyls and that hippopotamids are their extant sister group.

The hippopotamid hypothesis states that the fossil sister group of cetaceans is an artiodactyl in the lineage of hippopotamids, but not necessarily a species classified in the hippopotamid family. Hippopotamids are known exclusively from the Old World and go back only to the Miocene⁴², whereas cetaceans had already diverged in the Eocene of Asia^{2,12,43}. This implies that the likely cetacean sister group is an Old World artiodactyl of Eocene or older age¹¹. Furthermore, it should be realized that the mesonychian and hippopotamid hypotheses are not mutually exclusive. These hypotheses are in agreement if mesonychians are the cetacean sister group and this entire clade (called Cete) is the sister to the hippopotamid lineage.

In order to test whether the mesonychian hypothesis is robust against the addition of the new morphological data, we analysed a data matrix of dental, cranial and postcranial characters, and included a variety of mesonychians and artiodactyls (Fig. 4). Most importantly, we used our new fossil evidence to score pakicetids and improved our scoring of ambuloctids¹⁹. Our analysis (Fig. 4) supports Cetartiodactyla (the clade that includes Cetacea and Artiodactyla, to the exclusion of mesonychians), but not Cete (monophyletic Cetacea plus Mesonychia). There is strong bootstrap support (77%) for Cetartiodactyla, and we reject the mesonychian hypothesis of cetacean relations. However, our analysis does not support the hippopotamid hypothesis either, because we recovered sister group relations between Cetacea and monophyletic Artiodactyla.

Traditionally, the morphology of the ankle has been used to define artiodactyls (in a character-based definition^{23,25,27,42}). Our new fossils show that these defining characteristics do not only occur in all artiodactyls, but are also present in basal cetaceans. These ankle characters (deeply grooved proximal trochlea, dorso-plantar rotation plane of trochleated head, rectangular and wide sustentacular facet, flat and lateral ectal facet, elongate and oblique calcaneo-cuboid joint) have high consistency indices (1.0). It is now clear that they support the cetartiodactyl node (or the Cetartiodactyla + *Andrewsarchus* node, since no tarsus is known for this genus) in all most-parsimonious trees. Importantly, in none of the most-parsimonious trees was the mesonychian tarsus interpreted as

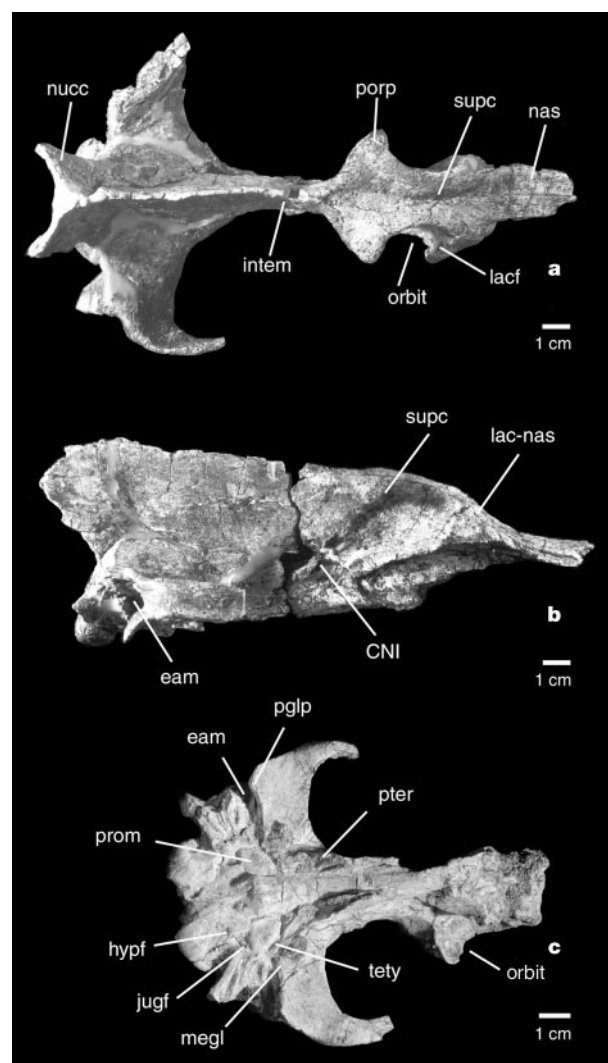


Figure 3 Skulls of the pakicetids *Pakicetus* (H-GSP 96231) in dorsal (a) and lateral (b) view and *Ichthyolestes* (H-GSP 98134) in ventral view (c). Abbreviations: CNI, cranial nerve I, endocast; eam, external auditory meatus; hypf, hypoglossal foramen; intem, intertemporal region; jugf, jugular foramen; megl, medial part of glenoid fossa; nas, nasal bone; nucc, nuchal crest; lacf, lacrimal foramen; lac-nas, lacrimo-nasal suture; porp, postorbital process; pglp, postglenoid process; prom, promontorium; pter, pterygoid process; supc, supraorbital canal; tety, tensor tympani fossa.

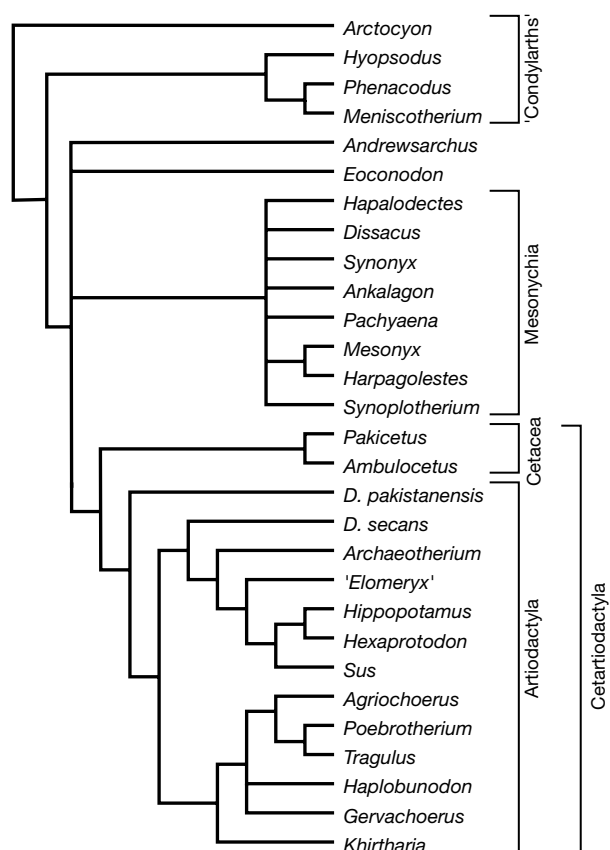


Figure 4 Phylogenetic relations of cetaceans to artiodactyls, mesonychians and primitive ungulates. Strict consensus cladogram of 38 most-parsimonious trees (see Methods for details).

a reversal from an artiodactyl-like morphology (the evolutionary model under which both hippopotamid and mesonychian hypotheses could be true).

The Artiodactyla node in our analysis is inconsistent with the hippopotamid hypothesis for cetacean relations, but is supported by fewer characters and has a lower bootstrap value (48%) than the Cetartiodactyla node. The artiodactyl node does receive support from the unique morphology of the cetacean fourth deciduous premolar (with consistency index, minimum possible number of changes/actual number of changes, of 1)^{37,44}. It is further supported by the reversal of a number of dental apomorphies, such as the large molar metacone and trigon basin, the presence and equal size of paracune and metacone, and a relatively low trigonid. Our analysis implies that the relatively primitive dental morphology of archaic artiodactyls is either a reversal (from a more mesonychian-like morphology) or that mesonychians and cetaceans evolved dental similarities independently. This bears out the prediction that widespread homoplasy occurred in one organ system in the early evolution of the clades in question⁴⁵. Our data suggest that the dentition, not the tarsus, was this organ system. This interpretation is consistent with a previous analysis that concluded that dental data are the primary source for the discrepancy between mesonychian and hippopotamid hypotheses⁴⁶, and that non-dental morphological data are more compatible with the hippopotamid hypothesis.

We suggest that the key to testing the hippopotamid hypothesis lies in the study of the more than ten families of early and middle Eocene artiodactyls from the Old World. Such a study may uncover the Eocene roots of the hippopotamid lineage, and its relation to cetaceans.

Note added in proof: Close cetacean-artiodactyl relations are also implied by protocetid fossils in an upcoming paper⁴⁸. □

Methods

Collection and identification

All pakicetid fossils were found at the early Eocene Howard University/Geological Survey of Pakistan (H-GSP) Locality 62 in the Ganda Kas Area of the Kala Chitta Hills (Punjab, Pakistan)³⁰. This locality outcrops over approximately 20 m² and contains a rich assemblage of isolated specimens with very low diversity. To date, 105 positively identifiable dental specimens have been recovered; 61% of these are pakicetid whales, 11% are anthracobunid proboscideans, 14% are raoellid artiodactyls, and the remainder is made up of small mammals (rodents, insectivores and mouse-sized marsupials³⁰). On the basis of these proportions, cetaceans can be expected to dominate the non-dental remains of this assemblage too. The largest cetaceans (*Pakicetus*, *Nalacetus*) are similar in size to the anthracobunids, but are easily differentiated on the basis of the distinctive postcranial morphology of the latter (which is known from several partial associated skeletons from the Ganda Kas Area). The smallest cetacean (*Ichthyolestes pinfoldi*) is approximately 174% as large (in linear dimensions) as the only known artiodactyl (the raoellid *Khirtharia dayi*) at Locality 62, and bones of these genera are thus not easily confused.

Initial identifications were based on size and anatomical fit between elements, and comparisons to *Ambulocetus*. To test these identifications, we analysed some bones isotopically. This method is destructive, so we limited the number of samples for this test. Isotopically, cetaceans are distinctive; their $\delta^{13}\text{C}$ values of enamel, dentin, and (mandibular) bone range between -11.9 and -13.6 ($n = 11$), whereas the same tissues for raoellid artiodactyls ($n = 5$) and anthracobunids ($n = 4$) range between -8.4 and -11.5 . Thus, among the large mammals at this locality, cetaceans have distinct isotopic values. This method allowed us to isotopically confirm the identity of a cervical vertebra of *Pakicetus* (H-GSP 92082, $\delta^{13}\text{C} = -13.9$), a lumbar vertebra *Pakicetus* (H-GSP 96284, $\delta^{13}\text{C} = -15.0$), a sacrum of *Pakicetus* (H-GSP 30251, $\delta^{13}\text{C} = -13.4$), a caudal vertebra of *Pakicetus* (H-GSP 96422, $\delta^{13}\text{C} = -14.8$), a humerus of *Pakicetus* (H-GSP 92042, $\delta^{13}\text{C} = -14.3$), a humerus of *Ichthyolestes* (H-GSP 96247, $\delta^{13}\text{C} = -13.0$), an innominate of *Pakicetus* (H-GSP 30279, $\delta^{13}\text{C} = -13.7$), an innominate of *Ichthyolestes* (H-GSP 30390, $\delta^{13}\text{C} = -13.0$), a tibia of *Pakicetus* (H-GSP 30315, $\delta^{13}\text{C} = -14.3$), an astragalus of *Ichthyolestes* (H-GSP 97001, $\delta^{13}\text{C} = -13.3$), two astragali of *Ichthyolestes* (H-GSP 98148, $\delta^{13}\text{C} = -13.4$; H-GSP 98149, $\delta^{13}\text{C} = -13.5$), a calcaneum of *Pakicetus* (H-GSP 96420, $\delta^{13}\text{C} = -12.4$), and a metatarsal of *Pakicetus* (H-GSP 30417, $\delta^{13}\text{C} = -14.0$).

There are several distinctive shape differences in the postcranial bones between *Ichthyolestes* and the larger pakicetids. Although *Nalacetus* differs from *Pakicetus* in dental and bullar³ morphology, no shape differences were found in this size cohort in the postcranial skeleton. It is thus likely that some of the bones referred to *Pakicetus* above pertain to *Nalacetus*, but this does not alter any of our conclusions. Our present study also shows that an astragalus attributed to a pakicetid^{24,25} was misattributed.

Cladistic analysis

In order to test the mesonychian and hippopotamid hypotheses, we chose in-group taxa that both sample ungulate diversity and remain pertinent to the question of sister group relations to Cetacea. We chose four archaic ungulates to root the tree in the ungulate radiation with *Arctocyon* as outgroup, a practice followed by many other studies^{4,44,46}. We limited Cetacea to two taxa (pakicetids, *Ambulocetus*) because the identity and basal status of these taxa are generally accepted^{1,45} and the inclusion of more taxa would make the analysis unwieldy. We included a large sample of mesonychians, following a previous study⁴. We also included 13 artiodactyls, sampling across the breadth of the modern suborders (Suina, Tylopoda and Ruminantia), but focusing on hippopotamids, their relatives, and Eocene artiodactyls. Following a previous study⁴² we suspected that the genus *Diacodexis* was paraphyletic. Therefore, we treated its North American *D. secans* and Asian *D. pakistanensis* as separate clades. We also added three families of Eocene Old World artiodactyls: cebochoerids, haplobunodontids, and raoellids. An explicit phylogenetic analysis of artiodactyls⁴² found raoellids near the base of a radiation of bunodont artiodactyls (including hippopotamids), and cebochoerids to be in an unresolved trichotomy with hippopotamids as part of the second clade, and anthracotheriids as the third. Anthracotheriids (represented in this analysis by *Elomeryx*⁴) are commonly considered to be closely related to hippopotamids, and haplobunodontids are traditionally included in the stem group of anthracotheriids. European haplobunodontids are thought to be middle Eocene migrants from Asia, making them possible candidates for the cetacean sister group under the hippopotamid hypothesis. Only three families of artiodactyls are known from Eocene Indo-Pakistan, the probable birthplace of cetaceans. We include all three in our analysis: raoellids, dichobunids (*D. pakistanensis*) and anthracotheriids.

Our matrix contains 29 taxa and 105 characters, 95 of which are parsimony informative and eight of which are ordered. Although gaps and missing characters represent different types of data, they are scored the same. We analysed the matrix using heuristic search algorithms in PAUP 4.0b8 software⁴⁷. ACCTRAN and DELTRAN optimizations were performed to investigate character transformation and estimated bootstrap values were calculated. Our maximum parsimony analysis produced 38 trees with a length of 281, consistency index of 0.39, and retention index of 0.61. A strict consensus cladogram of these trees is provided in Fig. 4. Character descriptions and scores, their sources, our matrix and our analyses are presented as Supplementary Information. Our heuristic search recovered a second island of trees at 282 steps. There are 38 trees with this tree length and both the Cetartiodactyla and Artiodactyla nodes were maintained in all trees.

Our results are robust against addition of *Hyopsodus* as an outgroup, to exclusion of *Menisotherium* and *Phenacodus* (which are dentally derived), and to the exclusion of *Eoconodon* and *Andrewsarchus* (which are in important positions in our cladograms, but are poorly known). The latter two taxa are the primary reason why character support for the Cetartiodactyla node varies with ACCTRAN and DELTRAN optimizations. ACCTRAN optimization shifts several important character changes from the Cetartiodactyla node to the node of Cetartiodactyla + *Andrewsarchus*.

Received 10 August; accepted 28 August 2001.

- Hulbert, R. C. Jr, Petkewich, R. M., Bishop, G. A., Bukry, D. & Aleshire, D. P. A new middle Eocene protocetid whale (Mammalia: Cetacea: Archaeoceti) and associated biota from Georgia. *J. Paleontol.* **72**, 907–927 (1998).
- Thewissen, J. G. M. & Hussain, S. T. Systematic review of the Pakicetidae, early and middle Eocene Cetacea (Mammalia) from Pakistan and India. *Bull. Carnegie Mus. Nat. Hist.* **34**, 220–238 (1998).
- Luo, Z. & Gingerich, P. D. Terrestrial Mesonychia to aquatic Cetacea: transformation of the basicranium and evolution of hearing in whales. *Univ. Michigan Pap. Paleontol.* **31**, 1–98 (1999).
- O'Leary, M. A. & Geisler, J. H. The position of Cetacea within Mammalia: phylogenetic analysis of morphological data from extinct and extant taxa. *Syst. Biol.* **48**, 455–490 (1999).
- Uhen, M. D. New species of protocetid archaeocete whale, *Eocetus wardii* (Mammalia, Cetacea) from the middle Eocene of North Carolina. *J. Paleont.* **73**, 512–528 (1999).
- Luo, Z. in *The Emergence of Whales: Evolutionary Patterns in the Origin of Cetacea* (ed. Thewissen, J. G. M.) 269–301 (Plenum, New York, 1998).
- Thewissen, J. G. M., Madar, S. I. & Arif, M. Fossil evidence for the origin of aquatic locomotion in archaeocete whales. *Science* **263**, 210–212 (1994).
- Thewissen, J. G. M., Madar, S. I. & Hussain, S. T. *Ambulocetus natans*, an Eocene cetacean (Mammalia) from Pakistan. *Courier Forschungsinstitut Senckenberg* **191**, 1–86 (1996).
- Thewissen, J. G. M. & Fish, F. E. Locomotor evolution in the earliest cetaceans: functional model, modern analogues, and paleontological evidence. *Paleobiology* **23**, 482–490 (1997).
- Milinkovitch, M. C. & Thewissen, J. G. M. Even-toed fingerprints on whale ancestry. *Nature* **388**, 622–624 (1997).
- Luo, Z. In search of whales' sisters. *Nature* **404**, 235–237 (2000).
- Gingerich, P. D., Russell, D. E. & Shah, S. M. I. Origin of whales in epicontinental remnant seas: new evidence from the early Eocene of Pakistan. *Science* **220**, 403–406 (1983).
- Kellogg, R. A. *Review of the Archaeoceti* (Carnegie Inst. Washington, Washington, 1936).
- Gingerich, P. D., Smith, B. H. & Simons, E. L. Hind limbs of Eocene *Basilosaurus*: evidence of feet in whales. *Science* **249**, 154–157 (1990).
- Uhen, M. D. in *The Emergence of Whales: Evolutionary Patterns in the Origin of Cetacea* (ed. Thewissen, J. G. M.) 29–61 (Plenum, New York, 1998).
- Bajpai, S. & Thewissen, J. G. M. A new diminutive Eocene whale from Kachchh (Gujarat, India) and its implications for locomotor evolution of cetaceans. *Curr. Sci.* **79**, 1478–1482 (2000).
- Zhou, X., Sanders, W. J. & Gingerich, P. D. Functional and behavioral implications of vertebral structure in *Pachyaena ossifraga* (Mammalia, Mesonychia). *Contrib. Mus. Paleontol. Univ. Michigan* **28**, 289–319 (1992).
- Thewissen, J. G. M. & Hussain, S. T. Postcranial osteology of the most primitive artiodactyl *Diacodexis pakistanensis* (Dichobunidae). *Anat. Histol. Embryol.* **19**, 37–48 (1990).
- Madar, S. I., Thewissen, J. G. M. & Hussain, S. T. Additional holotype remains of *Ambulocetus natans* (Cetacea, Ambulocetidae), and their implications for locomotion in early whales. *J. Vertebr. Paleont.* (in the press).
- Howell, A. B. *Aquatic Mammals: Their Adaptations to Life in the Water* (Charles Thomas, Baltimore, 1930).
- Howell, A. B. *Speed in Animals: Their Specializations for Running and Leaping* (Univ. Chicago Press, Chicago, 1944).
- King, J. E. *Seals of the World* (Cornell Univ. Press, Ithaca, 1991).
- Shaeffer, B. Notes on the origin and function of the artiodactyl tarsus. *Am. Mus. Novit.* **1356**, 1–26 (1947).
- Thewissen, J. G. M. & Madar, S. I. Ankle morphology of the earliest cetaceans and its implications for the phylogenetic relations among ungulates. *Syst. Biol.* **48**, 21–30 (1999).
- Vaughan, T. A. *Mammalogy* (Saunders College, New York, 1986).
- Bajpai, S. & Thewissen, J. G. M. in *The Emergence of Whales: Evolutionary Patterns in the Origin of Cetacea* (ed. Thewissen, J. G. M.) 213–233 (Plenum, New York, 1998).
- Weber, M. *Die Säugetiere* (Gustav Fischer, Jena, 1928).
- Gingerich, P. D., Raza, M. S., Arif, M., Anwar, M. & Zhou, X. New whale from the Eocene of Pakistan and the origin of cetacean swimming. *Nature* **368**, 844–847 (1994).
- O'Leary, M. A. & Uhen, M. D. The time of origin of whales and the role of behavioral changes in the terrestrial-aquatic transition. *Paleobiology* **25**, 534–556 (1999).
- Thewissen, J. G. M., Williams, E. M. & Hussain, S. M. Eocene mammal faunas from northern Indo-Pakistan. *J. Vertebr. Paleont.* **21**, 347–366 (2001).
- Thewissen, J. G. M. & Hussain, S. T. Origin of underwater hearing in whales. *Nature* **361**, 444–445 (1993).
- Hemilä, S., Nummela, S. & Reuter, T. A model of the odontocete middle ear. *Hearing Res.* **133**, 82–97 (1999).
- Rado, R., Himelfarb, M., Arensberg, B., Terkel, J. & Wollberg, Z. Are seismic communication signals transmitted by bone conduction in the blind mole rat? *Hearing Res.* **41**, 23–30 (1989).
- Lenhardt, M. L. Bone conduction hearing in turtles. *J. Audit. Res.* **22**, 153–160 (1982).
- Geisler, J. H. & Luo, Z. in *The Emergence of Whales: Evolutionary Patterns in the Origin of Cetacea* (ed. Thewissen, J. G. M.) 163–212 (Plenum, New York, 1998).
- O'Leary, M. A. in *The Emergence of Whales: Evolutionary Patterns in the Origin of Cetacea* (ed. Thewissen, J. G. M.) 133–161 (Plenum, New York, 1998).
- Lockett, W. P. & Hong, N. Phylogenetic relationships between the orders Artiodactyla and Cetacea: a combined assessment of morphological and molecular evidence. *J. Mamm. Evol.* **5**, 127–182 (1998).
- Shimamura, M. *et al.* Molecular evidence from retrotransposons that whales form a clade within even-toed ungulates. *Nature* **388**, 666–670 (1997).
- Gatesy, J. in *The Emergence of Whales: Evolutionary Patterns in the Origin of Cetacea* (ed. Thewissen, J. G. M.) 63–111 (Plenum, New York, 1998).
- Milinkovitch, M. C., Bérubé, M. & Palsbøll, P. J. in *The Emergence of Whales: Evolutionary Patterns in the Origin of Cetacea* (ed. Thewissen, J. G. M.) 113–131 (Plenum, New York, 1998).
- Nikaido, M., Rooney, A. P. & Okada, N. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: hippopotamuses are the closest extant relatives of whales. *Proc. Natl Acad. Sci. USA* **96**, 10261–10266 (1999).
- Gentry, A. W. & Hooker, J. J. in *The Phylogeny and Classification of the Tetrapods, Vol. 2: Mammals* (ed. Benton, M. J.) 235–272 (Clarendon, Oxford, 1988).
- Bajpai, S. & Gingerich, P. D. A new Eocene archaeocete (Mammalia, Cetacea) from India and the time of origin of whales. *Proc. Natl Acad. Sci. USA* **95**, 15464–15468 (1998).
- Thewissen, J. G. M. Phylogenetic aspects of cetacean origins: a morphological perspective. *J. Mamm. Evol.* **2**, 157–184 (1994).
- Thewissen, J. G. M., Madar, S. I. & Hussain, S. T. Whale ankles and evolutionary relationships. *Nature* **395**, 452 (1998).
- Naylor, G. J. P. & Adams, D. C. Are the fossil data really at odds with the molecular data? Morphological evidence for Cetartiodactyla phylogeny reexamined. *Syst. Biol.* **50**, 444–453 (2001).
- Swofford, D. L. *PAUP 4.0b8 Phylogenetic Analysis Using Parsimony (and Other Methods)* (Sinauer, Sunderland, Massachusetts, 1998).
- Gingerich, P. D., Haq, M.-u., Zalmout, I., Khan, I. H. & Malkani, M. S. Origin of whales from early artiodactyls: hands and feet of Eocene Protocetidae from Pakistan. *Science* (in the press).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank M. Arif, S. Bajpai, J. Erfurt, A. Friscia, M. Hellmund, S. Madar, M. Raza, J. Quade and the Geological Survey of Pakistan for assistance in field work, access to collections and laboratories, and/or discussions. M. Tomasko prepared Fig. 1. Funding for this research was provided by the National Science Foundation (EAR 9902830).

Correspondence and requests for materials should be directed to J.G.M.T. (e-mail: thewisse@neucom.edu).

Microscopic electronic inhomogeneity in the high- T_c superconductor $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$

S. H. Pan*, J. P. O'Neal*, R. L. Badzey*, C. Chamon*, H. Ding†, J. R. Engelbrecht†, Z. Wang†, H. Eisaki‡§, S. Uchida‡, A. K. Gupta||, K.-W. Ng||, E. W. Hudson¶§, K. M. Lang¶ & J. C. Davis¶

* Department of Physics, Boston University, Boston, Massachusetts 02215, USA

† Department of Physics, Boston College, Chestnut Hill, Massachusetts 02467, USA

‡ Department of Superconductivity, University of Tokyo, Yayoi, 2-11-16 Bunkyo-ku, Tokyo 113-8656, Japan

|| Department of Physics and Astronomy, University of Kentucky, Lexington, Kentucky 40506-0055, USA

¶ Department of Physics, University of California, Berkeley, California 94720, USA

The parent compounds of the copper oxide high-transition-temperature (high- T_c) superconductors are unusual insulators (so-called Mott insulators). Superconductivity arises when they are 'doped' away from stoichiometry¹. For the compound $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$, doping is achieved by adding extra oxygen atoms, which introduce positive charge carriers ('holes') into the CuO_2 planes where the superconductivity is believed to originate. Aside from providing the charge carriers, the role of the oxygen dopants is not well understood, nor is it clear how the charge carriers are distributed on the planes. Many models of high- T_c superconductivity accordingly assume that the introduced carriers are distributed uniformly, leading to an electronically homogeneous system as in ordinary metals. Here we report the presence of an electronic inhomogeneity in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ on the basis of observations using scanning tunnelling microscopy and spectroscopy. The inhomogeneity is manifested as spatial variations in both the local density of states spectrum and the superconducting energy gap. These variations are correlated spatially and vary on the surprisingly short length scale of ~ 14 Å. Our analysis suggests that this inhomogeneity is a consequence of proximity to a Mott insulator resulting in poor screening of the charge potentials associated with the oxygen ions left in the BiO plane after doping, and is indicative of the local nature of the superconducting state.

We have carried out extensive low-temperature scanning tunnelling microscopy/spectroscopy (STM/S) studies on optimally doped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$. The technical details of the experiments have been reported previously^{2,3}. Various single crystal samples fabricated with different techniques (directional-flux solidification or floating-zone) have been used to ensure the generality of the observations. Some crystals are nominally pure and some are deliberately doped with a very dilute concentration of impurity atoms (Zn or Ni). We have consistently observed a particular type of electronic inhomogeneity regardless of the differences between the individual crystals. This inhomogeneity is manifested as spatial variations in the local density of states (LDOS) spectrum, in the low-energy spectral weight, and in the magnitude of the superconducting energy gap.

As an example, Fig. 1a presents a topographic image obtained on a pure $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ crystal grown by the flux-solidification method. The image reveals the crystal structure of the BiO plane with atomic resolution accompanied by the well known incommensurate structural modulations.

In addition, an inhomogeneous background is also discernible. After removing the contrast associated with the above-mentioned topological structures by Fourier filtering, this inhomogeneous background is clearly visible. Assuming that the tunnelling matrix element has no spatial dependence, the resulting image shown in Fig. 1b can be approximated by a map displaying the variation of the integrated LDOS. With spatially resolved spectroscopy, we find that the local tunnelling spectrum also varies from location to location. We discover that the magnitude of the superconducting gap extracted from the local tunnelling spectrum varies spatially as well, instead of exhibiting the single-valued nominal gap of ~ 40 meV that is usually observed by tunnelling in crystals with optimal oxygen doping^{4,5}. Both the integrated LDOS and the superconducting gap magnitude vary on

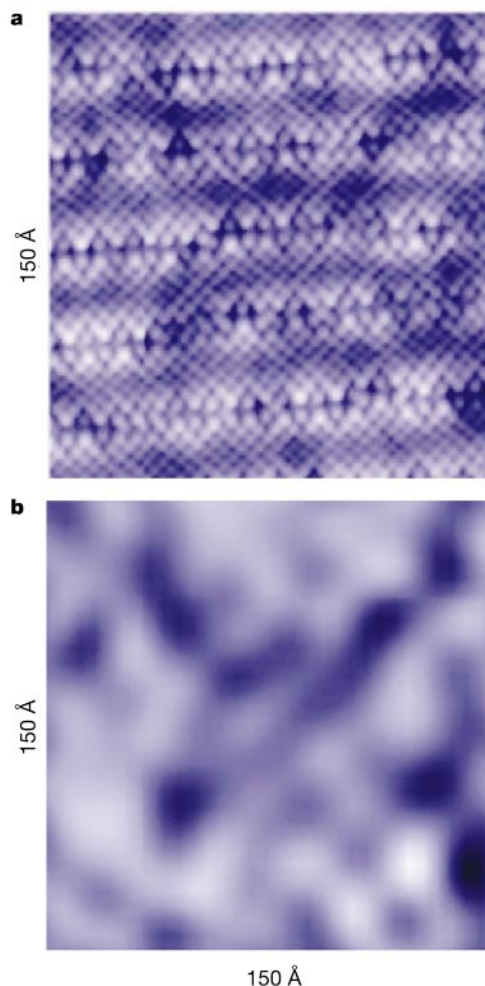


Figure 1 Topographic image and associated integrated LDOS map of an optimally oxygen-doped, nominally pure single crystal of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$. **a**, Constant-current-mode, topographic image (150×150 Å) of the surface BiO plane exposed after cleavage of the single crystal. In addition to the clear contrast due to the topological corrugations of the atomic structures and the well known incommensurate structural modulation, an inhomogeneous background is also visible. In constant-current mode, the tunnelling current is exponentially related to the tip-sample distance and is also proportionally related to the integrated LDOS. Therefore, the constant current topographic image provides the convolved information of both the topology and the LDOS of the crystal surface. **b**, To reveal the inhomogeneous background more clearly, we use Fourier filtering to remove the contrast due to the two well-ordered topological structures mentioned above. The variation of the integrated LDOS, which is seen as an inhomogeneous background in **a**, is now clearly displayed. A brighter colour represents a larger magnitude of the integrated LDOS.

§ Present addresses: Department of Applied Physics, Stanford University, Stanford, California 94305, USA (H.E.); and National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA (E.W.H.).

an apparently much shorter length scale than those in the phenomena observed in earlier experiments^{6,7}. These observations are surprising because in conventional Bardeen–Cooper–Schrieffer (BCS) superconductors (such as Nb) the integrated LDOS, the tunnelling spectrum, and the superconducting gap are spatially homogeneous.

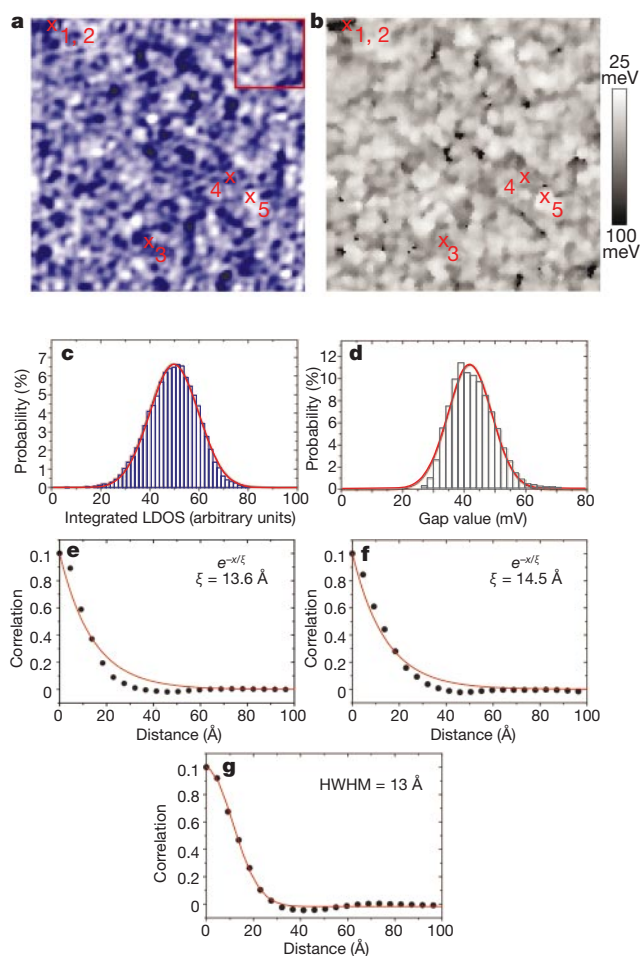


Figure 2 A comparison of an integrated LDOS map and its corresponding superconducting gap map, including their associated statistical results. **a**, A $600 \times 600 \text{ Å}$ LDOS map obtained with the same technique used in Fig. 1b on a single crystal $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ doped with a very dilute concentration of Zn atoms. (The Zn concentration measured by STM is 0.2%; ref. 2.) The crystal has a superconducting transition temperature of 84 K, with a transition width of 4 K. The red box in panel **a** frames a $150 \times 150 \text{ Å}$ region for easy comparison with Fig. 1b. It clearly displays an inhomogeneous structure similar to that observed in Zn-free crystals in Fig. 1b. **b**, Superconducting gap map, obtained simultaneously with the integrated LDOS map on the same location, showing the spatial variation of the superconducting energy gap. The local gap values are extracted from the corresponding local differential conductance spectra. We assigned reverse colour coding (higher intensity corresponds to a smaller gap magnitude) to the map so that it is easier to visualize its correlation with the integrated LDOS map. **c** and **d** are the histograms showing the statistical distributions of the integrated LDOS and the magnitude of the superconducting gap. Each of them exhibits a gaussian-like distribution (fitting function displayed in red). The fit of the gap distribution (42 meV mean; $\sim 20 \text{ meV}$ FWHM) shows it to be slightly skewed. **e** and **f** show the azimuthally averaged results of the two-dimensional auto-correlation analysis on the integrated LDOS map and superconducting gap map respectively. Fitting both data sets with a simple exponential decay model (red lines) results in decay lengths of $\sim 14 \text{ Å}$, demonstrating the short length scale of the variations. The imperfections of the fit imply that a more complex model might be needed. **g**, The superconducting gap is spatially correlated with the LDOS as characterized by this two-dimensional cross-correlation function. It has a pronounced centre peak that can be fitted with a gaussian function.

What is the origin of such an inhomogeneity? Is it intrinsic or caused by impurities such as crystal defects or excess elements that limit the sample quality? Figure 2 is an example of the STM/S results obtained on high quality, floating-zone-grown single crystals deliberately doped with a very dilute concentration (0.2%) of Zn atoms. The inhomogeneous background observed in these crystals (Fig. 2a) is of the same type as that seen in Fig. 1b. This suggests that the inhomogeneity does not originate from impurities. To support this point, we identified the locations of the Zn impurities from a zero bias conductance map, which is taken simultaneously with the LDOS map at the same location³. Using cross-correlation analysis, we found no correlation between the intensity of the integrated LDOS and the locations of the Zn impurities. Furthermore, the length scale of the spatial variation observed in the integrated LDOS map is much smaller than the average impurity spacing. Therefore, we conclude that the inhomogeneity observed in the integrated LDOS is not induced by impurities, but rather is intrinsic in nature.

Spatial variations of the tunnelling spectrum and of the superconducting gap, similar to those observed in the pure sample, are observed in this impurity-doped sample (Fig. 2b). Statistical analysis of Fig. 2a shows that the integrated LDOS has a gaussian distribution, displayed in Fig. 2c. Similarly, analysis of Fig. 2b shows that the gap ranges from 25 meV to 65 meV and exhibits a gaussian distribution (42 meV mean; $\sim 20 \text{ meV}$ full-width at half-maximum, FWHM) as shown in Fig. 2d. We note that the average gap value is very similar to the single value reported on optimally doped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ in earlier tunnelling measurements^{4,5}.

In addition to the similarity of their statistical distributions, a strong spatial correlation between the integrated LDOS and the superconducting gap can be seen readily by recognizing similar patterns in the two maps (Fig. 2a and b), which are obtained at the same location simultaneously. As illustrated in Fig. 2e–g, auto-correlation analysis on both maps shows similar decay lengths ξ of approximately 14 Å and cross-correlation shows a pronounced gaussian peak, confirming the local nature of the variations and the strong correlation of these two quantities. Furthermore, the line-shape of the local tunnelling spectrum also correlates with the integrated LDOS. As the gap and the integrated LDOS variations are correlated and the energy gap variation cannot be attributed to a tunnelling matrix effect, the inhomogeneity in the integrated LDOS we observe is most probably intrinsic to the electronic structure and not due to a spatially varying matrix element.

Spatial variation of the energy gap and its correlation with the LDOS can be seen in greater detail in Fig. 3. Clearly, the spectra obtained at points with larger integrated LDOS values exhibit higher differential conductance, smaller gap values, and sharper coherence peaks. More significantly, these STM/S results resemble previous tunnelling spectra obtained on samples with different oxygen doping concentrations^{4,5}—the STM/S spectra obtained at positions with higher values of integrated LDOS resemble characteristic spectra obtained on samples with higher oxygen doping concentrations. These observations naturally lead one to relate the magnitude of the integrated local DOS to the local oxygen doping concentration. On a macroscopic scale, such a correspondence is expected for a doped Mott insulator where doping adds spectral weight near the Fermi energy and provides the carriers necessary to transform the insulating compound into a superconductor, as observed previously by photoemission⁸. The idea of local doping concentration (LDC) extends this picture to microscopic scales. In the LDC framework, the charge potentials of oxygen dopants alter the local electronic environment provided that the screening of these potentials is weak on the scale of the inter-carrier distance. For doped Mott insulators, this condition is satisfied by a relatively large inter-carrier distance and strong electronic correlation. A detailed theoretical analysis supporting the LDC framework will be discussed elsewhere.

In Fig. 4, we present a scatter-plot of the magnitude of the energy

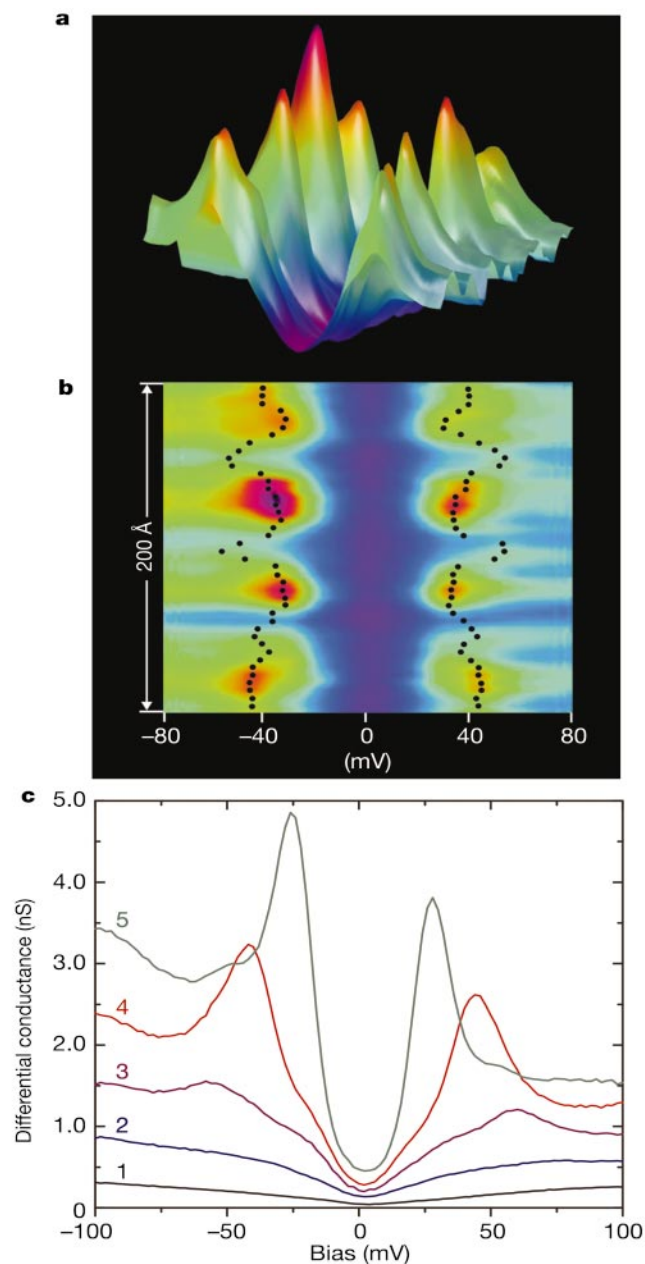


Figure 3 Spatial variation of the tunnelling differential conductance spectrum. To account for variations in the tunnelling junction, the spectra are normalized to a constant tip-sample separation. **a**, A three-dimensional rendering of the tunnelling spectra along a 200 Å line, showing detailed variations of the LDOS, the energy gap, and their correlation. We note that the coherence peak heights vary more dramatically than the gap widths. **b**, The same data as a bird's-eye view. To demonstrate the gap variation, the black dotted lines trace the positions of the coherence peaks. Together with the gap map of Fig. 2b, we can see that the gap varies less rapidly within a 'patch' of high DOS than at its edge. **c**, Five characteristic spectra taken at the positions correspondingly marked in both Fig. 2a and b showing the correlation between the superconducting gap and the integrated LDOS. Curves 1 and 2 are taken at nearby positions in a very dark area on the integrated LDOS map in Fig. 2a, where the integrated LDOS is very small. The low differential conductance and the absence of a superconducting gap are indicative of insulating behaviour. Curve 3 has a large gap of 65 meV, with low coherence peaks, resembling the spectral line-shape measured in oxygen-underdoped samples^{4,5}. The integrated value of the LDOS at the position for curve 3 is small but larger than those in curves 1 and 2, as it is in a slightly brighter area. Curve 4 comes from a still brighter area. It has a gap value of 40 meV, which is close to the mean value of the gap distribution. The coherence peaks are sharper and higher, resembling the spectral behaviour of a sample with optimal oxygen doping^{4,5}. Curve number 5, taken at the position with the highest integrated LDOS, has the smallest gap value of 25 meV with two very sharp coherence peaks, resembling the spectral behaviour of an oxygen over-doped crystal^{4,5}.

gap versus the value of the integrated LDOS. For comparison, the doping dependence of the gap value obtained by angle-resolved photoelectron spectroscopy (ARPES) is superimposed on the plot. We note that STM is a real-space, local probe, whereas ARPES resolves in momentum space but averages in real space over a macroscopic spot. The similarity in the linear behaviour of these two complementary measurements is remarkable, and provides further support for the concept of LDC in this strongly correlated system.

The existence of such microscopic inhomogeneities should have many important consequences on the quasiparticle properties that are accessible by macroscopic measurements. For example, unexpectedly broad peaks in an earlier neutron scattering measurement on $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ could possibly be explained by the presence of inhomogeneity in the bulk⁹. However, it is more important to compare our results to a complementary technique such as ARPES, which measures the same electronic excitations in momentum space under similar experimental conditions. In doing so, we find that taking a spatial average of the STM results yields the same maximum gap value and the same doping dependence of the gap as obtained by ARPES. In addition, the large intrinsic width (~ 20 meV) of the coherence peak measured by ARPES^{10,11} near the antinodes is consistent with the 20 meV FWHM distribution of the superconducting gap shown in Fig. 2d. In light of the electronic disorder observed by this STM experiment, the discrepancy between the nodal and anti-nodal quasiparticle mean free paths observed by ARPES can now be reconciled. Near the antinodes, gap disorder can significantly scatter the quasiparticles at the gap edge. Indeed, the peak in the momentum distribution curve (MDC) near the antinodes has a large width of $\sim 0.08 \text{ \AA}^{-1}$ (ref. 12), corresponding to a short mean free path of $\sim 25 \text{ \AA}$. In the nodal direction, however, ARPES measures a mean free path of $\sim 100 \text{ \AA}$ (refs 13 and 14). The quasiparticles of the *d*-wave superconductor in this direction are less affected by the gap disorder since the latter amounts to velocity disorder in the dispersion, which is not effective at scattering Dirac particles. Thus, the mean free path of the nodal quasiparticles will be

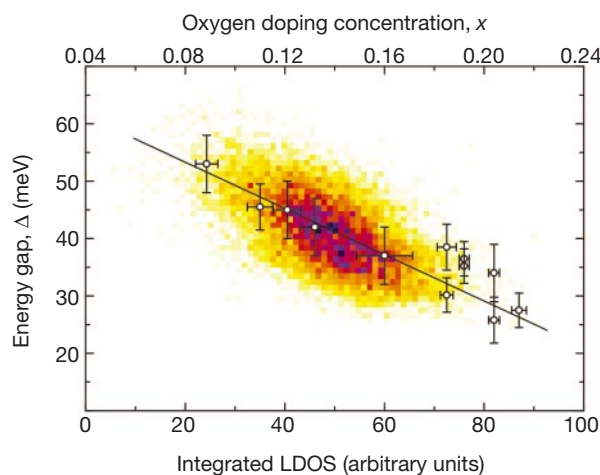


Figure 4 A scatter plot of the superconducting gap versus integrated LDOS. The 16,384 data points of the corresponding integrated LDOS map and the gap map (Fig. 2a, b) are plotted here in colour. A darker colour represents a larger number of data points with the same integrated LDOS and the same gap magnitude. For comparison, a set of data from ARPES measurements on various $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ single crystals with different oxygen doping concentrations is superimposed onto the plot. The gap values of ARPES data are the maximum values of the *d*-wave gap at $(\pi, 0)$ of the Brillouin zone. They are presented as open circles in the plot and the corresponding doping concentration scale is placed on top of the plot. The solid black line is the linear fit to both the STM and the ARPES data.

limited primarily by elastic scattering from the potential disorder. Given that the oxygen dopants responsible for this potential disorder do not reside in the CuO_2 plane, scattering is relatively weak. This can result in a longer mean free path and the measured 100 Å length scale is therefore quite reasonable. As scattering into all angles is involved in ARPES measurements, whereas transport is dominated by large-angle scattering of the nodal quasiparticles, our results may also reconcile the discrepancy between the mean free path measured by transport¹⁵ as compared to that measured by ARPES. Taking 14 Å as the length scale over which the disorder potential varies significantly, we deduce that the dominant elastic scattering process is limited by the wavevector $q = 1/14 \text{ Å}^{-1}$. The scattering angle, θ , given by $\sin(\theta/2) = q/2k_F$, where k_F is the Fermi wavevector, is indeed quite small at about 5° , which provides a possible explanation as to why the transport mean free path can be much longer than that measured by ARPES.

Discussion of our observations can also be extended to more fundamental issues, such as the coherence of the superconducting state. The coexistence of a high superconducting transition temperature with such a microscopic inhomogeneity implies that the superconducting coherence length is shorter than the mean free path. Our measured gap correlation length, $\xi \approx 14 \text{ Å}$, sets the length scale for the superconducting pair size in optimally doped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$. By evaluating the BCS expression $\xi_0 = \hbar v_F / \pi \Delta$, taking $\hbar v_F = 1.6 \text{ eV Å}$ from band dispersion near the nodes^{14,16} and the averaged gap at optimal doping as $\Delta = 0.04 \text{ eV}$, we obtain $\xi_0 \approx 13 \text{ Å}$, which is in good agreement with the correlation decay length ξ obtained from our experiment. Yet ξ appears to be shorter than the experimental in-plane superconducting coherence length $\xi_{ab} \approx 22\text{--}27 \text{ Å}$ (refs 17–19). In contrast to conventional BCS superconductors, it is conceivable that the amplitude and phase coherence in high- T_c superconductors have different length scales, because the ratio $R = 2\Delta/k_B T_c$ is no longer a constant. Recent ARPES measurements¹¹ suggest that $R \propto 1/x$. Thus we may expect the superconducting phase coherence length to be determined by $\hbar v_F / k_B T_c$, which scales as $1/x$ on the underdoped side. An extension of our correlation length and vortex core-size measurements to underdoped samples with various doping concentrations will perhaps distinguish the two length scales because they may have different doping (x) dependences.

The observation of microscopic spatial variations in both the carrier density and the superconducting gap, and the strong correlation between these variations, reveals the local inhomogeneous charge environment in these materials, and its intimate relationship with superconductivity. Further exploration of this frontier may lead to a greater understanding of how high- T_c superconductivity arises from doping a Mott insulator. □

Received 23 January; accepted 18 July 2001.

- Anderson, P. W. *The Theory of Superconductivity in the High- T_c Cuprates* (Princeton Univ. Press, Princeton, New Jersey, 1997).
- Hudson, E. W., Pan, S. H., Gupta, A. K., Ng, K.-W. & Davis, J. C. Atomic-scale quasiparticle scattering resonances in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Science* **285**, 88–91 (1999).
- Pan, S. H. *et al.* Imaging the effects of individual zinc impurity atoms on superconductivity in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Nature* **403**, 746–750 (2000).
- Renner, Ch. *et al.* Pseudogap precursor of the superconducting gap in under- and overdoped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. Lett.* **80**, 149–152 (1998).
- Miyakawa, N. *et al.* Predominantly superconducting origin of large energy gaps in underdoped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ from tunneling spectroscopy. *Phys. Rev. Lett.* **83**, 1018–1021 (1999).
- Hasegawa, H., Ikuta, H. & Kitazawa, K. in *Physical Properties of High Temperature Superconductors III* (ed. Ginsberg, D. M.) Ch. 7 (World Scientific, Singapore, 1992).
- Wolf, E. L., Chang, A., Rong, Z. Y., Ivanchenko, Yu. M. & Lu, F. Direct STM mapping of the superconducting energy gap in single crystal $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *J. Superconductivity* **70**, 355–360 (1994).
- Ino, A. *et al.* Doping dependent density of states and pseudogap behavior in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. *Phys. Rev. Lett.* **81**, 2124–2127 (1998).
- Fong, H. F. *et al.* Neutron scattering from magnetic excitations in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Nature* **398**, 588–591 (1999).
- Feng, D. L. *et al.* Signature of superfluid density in the single-particle excitation spectrum of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Science* **289**, 277–281 (2000).
- Dings, H. *et al.* Coherent quasiparticle weight and its connection to high- T_c superconductivity from angle-resolved photoemission. Preprint cond-mat/0006143 at (<http://xxx.lanl.gov>) (2000).

- Valla, T. *et al.* Temperature dependent scattering rates at the Fermi surface of optimally doped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. Lett.* **85**, 828–831 (2000).
- Valla, T. *et al.* Evidence for quantum critical behavior in the optimally doped cuprate $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Science* **285**, 2110–2113 (1999).
- Kaminski, A. *et al.* Quasiparticles in the superconducting state of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. Lett.* **84**, 1788–1791 (2000).
- Zhang, Y. *et al.* Giant enhancement of the thermal Hall conductivity κ_{xy} in the superconductor $\text{YBa}_2\text{Cu}_3\text{O}_7$. Preprint cond-mat/0008140 at (<http://xxx.lanl.gov>) (2000).
- Bogdanov, P. V. *et al.* Evidence for an energy scale for quasiparticle dispersion in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. Lett.* **85**, 2581–2584 (2000).
- Palstra, T. T. M. *et al.* Angular dependence of the upper critical field of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. B* **38**, 5102–5105 (1988).
- Renner, Ch., Revaz, B., Kadowaki, K., Maggio-Aprile, I. & Fischer, Ø. Observation of the low temperature pseudogap in the vortex cores of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. Lett.* **80**, 3606–3609 (1998).
- Pan, S. H. *et al.* STM studies of the electronic structure of vortex cores in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. Lett.* **85**, 1536–1539 (2000).

Acknowledgements

We acknowledge P. A. Lee and E. W. Plummer for their comments. We also thank P. W. Anderson, A. Balatsky, D. A. Bonn, A. Castro-Neto, E. Carlson, M. Franz, L. H. Greene, X. Hu, T. Imai, B. Keimer, S. A. Kivelson, K. Kitazawa, R. B. Laughlin, D.-H. Lee, A. H. MacDonald, A. Millis, N. P. Ong, Z.-X. Shen, H.-J. Tao, X.-G. Wen, Z.-Y. Weng, N.-C. Yeh, G.-M. Zhang and Z.-X. Zhong for helpful discussions. This work was supported by the NSF, the DOE, the Sloan Research Fellowship, the Research Corporation, the Miller Institute for Basic Research and a Grant-in-Aid for Scientific Research on Priority Area and a COE Grant from the Ministry of Education, Japan.

Correspondence and requests for materials should be addressed to S.H.P. (e-mail: shpan@bu.edu).

Friction and fracture

Eric Gerde & M. Marder

Computational and Applied Mathematics and Center for Nonlinear Dynamics, The University of Texas at Austin, Austin, Texas 78712, USA

Consider a block placed on a table and pushed sideways until it begins to slide. Amontons and Coulomb found that the force required to initiate sliding is proportional to the weight of the block (the constant of proportionality being the static coefficient of friction), but independent of the area of contact¹. This is commonly explained by asserting that, owing to the presence of asperities on the two surfaces, the actual area in physical contact is much smaller than it seems, and grows in proportion to the applied compressive force². Here we present an alternative picture of the static friction coefficient, which starts with an atomic description of surfaces in contact and then employs a multiscale analysis technique to describe how sliding occurs for large objects. We demonstrate the existence of self-healing cracks^{3–4} that have been postulated to solve geophysical paradoxes about heat generated by earthquakes^{5–11,25–27}, and we show that, when such cracks are present at the atomic scale, they result in solids that slip in accord with Coulomb's law of friction. We expect that this mechanism for friction will be found to operate at many length scales, and that our approach for connecting atomic and continuum descriptions will enable more realistic first-principles calculations of friction coefficients.

The most intriguing fact about friction is that it is proportional to the force pushing two objects together, but independent of the area of contact. Without contesting the prevalence of materials where the conventional picture involving asperities applies, we wish to point out an alternative way of explaining the same basic fact. Suppose two solids are in close contact over some distance. Suppose further that a wave of separation, a self-healing crack, can run along the interface, like a bump on a rug, leading one solid to slip over the other (Fig. 1). And suppose finally that the condition for such self-

healing cracks to arise depends only upon the ratio of horizontal to compressive stress. Then sliding begins in accord with Coulomb's law, and is naturally independent of the area of contact.

We can demonstrate in a simple but realistic setting that these suppositions are correct. We begin by describing a calculation that employs standard techniques from continuum mechanics^{12–14}. Imagine a semi-infinite elastic solid atop a rigid substrate, impose vertical and horizontal stresses σ_{yy} and σ_{xx} far from the boundary, and suppose that there is a self-healing crack of length l , travelling at speed v , and causing slip of amount Δu (Fig. 1). For all values of these five parameters we can calculate stress and displacement fields everywhere in the elastic solid, including cases where σ_{yy} is compressive^{13,14}. However, there is a problem—all of these solutions have an unphysical feature. Sufficiently near each crack tip, the crack faces oscillate and pass through each other infinitely often. Perhaps none of the solutions can actually occur in a physical system. On the other hand, if the intersections occur on a scale much smaller than an interatomic spacing, is there really a problem? The possibility of solids sliding over one another as we describe hinges upon the existence of self-healing cracks. Finding explicit solutions in continuous media has not been hard, but it has been difficult to establish a consensus on whether or not the solutions are physically acceptable^{15,16}. The physical problem has been recast in a number of mathematical contexts. For example, one can investigate two surfaces that remain absolutely flat and in contact at all times, with a friction law allowing them to move relative to one another^{14,15}. However, so long as the problem is stated in the continuum, the essential difficulty persists.

A way to resolve the conceptual problems is to leave continuum

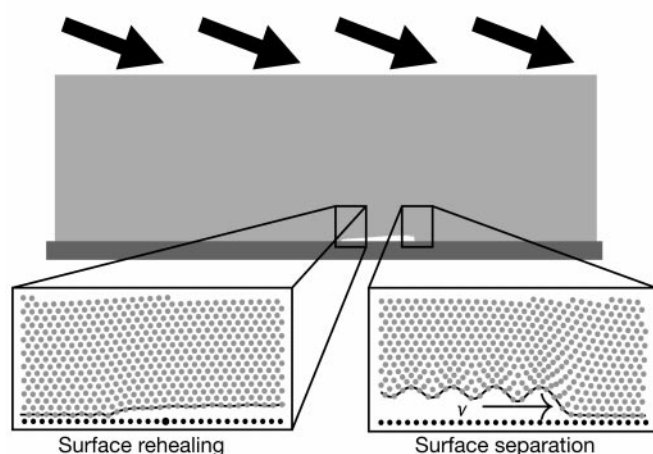


Figure 1 Numerical simulation of a self-healing crack travelling through a compressed strip. The block slips by a distance of three atoms over the lower surface when the self-healing crack moves from left to right. Atoms initially on a 400 by 1,200 triangular grid interact linearly with nearest neighbours. Interactions with the rigid substrate are governed by truncated hookean potentials with a bond rupture strain of 5%. Atoms from the upper block interact with any atom in the substrate that comes within range; this allows the upper surface to slip with respect to the lower and then heal. We impose fixed horizontal and vertical displacements on the top layer, and initiate dynamics with an artificially created crack. Creation of the moving crack proceeds in stages. At first, the upper layer of the system is in tension, and the tail end of the crack extends to the end of the system. Slowly the outer boundary is pushed into compression, at which point the left side of the crack closes up, and becomes self-healing. At the crack moves forward, we add new atoms to the right boundary, and remove atoms from the left. The system eventually converges to the steady-state shown here, with a self-healing crack of length $l = 168$, speed $v = 0.83c_s$, and surface slip $\Delta u = 3$. Distances are in units of interatomic spacing, and c_s is the shear wave speed of the material. Vertical displacements have been rescaled by 3 for visual clarity.

mechanics and pose the question about self-healing cracks at the atomic level. Cracks can be followed in atomic detail as they move through solids with arbitrarily large numbers of atoms¹⁷. We have extended the atomic-scale analytical techniques to include the case of interface fracture.

The context for our work is the following set of equations. Atomic equilibrium positions, $\mathbf{r}_{mn}$, lie on a triangular grid N atoms high with unit interatomic spacing:

$$\mathbf{r}_{mn} = \begin{bmatrix} m - n/2 \\ \sqrt{3}/2n \end{bmatrix}, \quad \begin{matrix} -\infty < m < \infty \\ 0 \leq n \leq N \end{matrix} \quad (1)$$

Displacements from equilibrium, $\mathbf{u}_{mn}(t)$, satisfy the steady-state equation, $\mathbf{u}_{mn}(t) = \mathbf{u}_{0n}(t - m/v)$, and the top layer is rigidly displaced, $\mathbf{u}_{mN}(t) = [\frac{D_x}{D_y}]$. The steady-state equations of motion are:

$$\begin{aligned} \ddot{\mathbf{u}}_{mn} + b\dot{\mathbf{u}}_{mn} = & \sum_{m',n' \text{ where } |\mathbf{d}|=1} (\mathbf{u}_{m'n'} - \mathbf{u}_{mn}) \cdot \mathbf{d} \mathbf{d} \\ & - \mathbf{u}_{mn} \cdot \mathbf{r}_{11} \mathbf{r}_{11} \theta(1/2 - n)\theta(m/v - t) \\ & - \mathbf{u}_{mn} \cdot \mathbf{r}_{01} \mathbf{r}_{01} \theta(1/2 - n)\theta(m/v - t + s) \end{aligned} \quad (2)$$

Here b is a small dissipation parameter, θ is the Heaviside function, and $\mathbf{d} = \mathbf{r}_{m'n'} - \mathbf{r}_{mn}$. The syncopation, s , designates the time difference between left and right bond ruptures for atoms along the interface. Finally, we require bonds to reach a critical strain, u_c , before rupture, and the newly created surface to remain separated thereafter. Apart from u_c , which sets an overall scale, in the limits $N \rightarrow \infty$, $b \rightarrow 0$ each microscopic fracture state is described by two parameters—its velocity v , and syncopation s . Methods for solving this problem generalize those described in ref. 18. The main new technical hurdle is that the equations reduce to a 2×2 Wiener–Hopf system. The system is approximated with Wilson's algorithm¹⁹ applied to an equivalent factorization problem on the unit circle²⁰ and solved.

The advantage of the atomic-scale analytical methods is that they can treat arbitrarily large numbers of atoms, and therefore lead to a connection with continuum analysis. However, they have an important limitation. The formal techniques only operate if bonds, once broken, never heal; they only treat semi-infinite cracks, and describe the forward tip of a self-healing crack, or its closing end, but not both at once.

The final stage of our analysis is to combine information from the continuum and the atomic solutions. The latter make it possible to determine when a crack tip is physically realizable and when it is not; the former make it possible to describe the self-healing process. The connection between continuous and atomic descriptions is created in the following way. Continuum calculations^{21,22} show that stress fields sufficiently close to the tip of an interface crack at the origin have an oscillating singularity where σ_{yy} takes the form:

$$\sigma_{yy}(x, 0) = \text{Re}[\tilde{K}x^{-1/2+i\epsilon}]/\sqrt{2\pi} \quad (3)$$

The real constant ϵ increases with the speed v of the crack, and the constant $\tilde{K}$ has both real and imaginary parts. Note that the sign of the stress oscillates as the crack tip is approached; the possibility of converting compressive stresses far away to tensile stresses near the tip helps explain in principle why self-healing cracks are possible.

Turning to an atomic point of view, we can find exactly the same universal form for the stress field, but it now emerges far from the tip of a semi-infinite crack. That is, by adopting a continuum perspective and 'zooming-in' on a crack tip, we find the same mathematical forms that are found by adopting an atomic perspective and 'zooming-out'. However, the atomic analysis has additional information to contribute. Once v and s are specified (see equation (2)), $\tilde{K}$ is completely determined.

Thus, the atomic analysis chooses from the five-dimensional

space of continuum solutions a subspace that is physically permitted. The precise states that are permitted depend upon details of the atomic model, but the elimination of most—but not all—continuum states should be general.

We are finally in a position to carry out an exhaustive search for self-healing cracks. The search proceeds in the following way. Fix v , l , and the total surface slip Δu . Next, fix the syncopation s_f for the forward crack tip (equation (2)) and use the atomic solutions to calculate the strength of the associated oscillating stress singularity $\bar{K}$. Inserting this value into the macroscopic solution, determine the stresses at infinity $\sigma_{xy}(\infty)$, $\sigma_{yy}(\infty)$. Next choose a syncopation s_b for the back of the crack. Again stresses at infinity are uniquely determined—but there is no reason they should be the same as the values dictated by the front. There are two parameters s_f and s_b that can be varied, leading to a finite number of cases where $\sigma_{xy}(\infty)$, $\sigma_{yy}(\infty)$ make both the front and back propagate at the prescribed speed. We typically find no solution, or one solution, on varying s_f

and s_b within the allowable range of syncopation values, roughly $|s| < 1/|v|$.

The results of this search are illustrated in Fig. 2. In Fig. 2a, we catalogue cracks moving between 50% and 80% of the shear wave speed, in velocity increments of 1%. Figure 2b focuses on the self-healing cracks that move at 80% of the shear wave speed. (We note that strains within the Earth, given by the ratio of a characteristic stress σ_{yy} to Young's modulus E , are of the order of 10^{-4} .) The lower boundary in Fig. 2a is roughly linear, and extends down to strains of the order of 10^{-5} . The complex structure of solutions for strains below 10^{-5} results from the existence of a minimum crack-propagation speed. Such minimum speeds are typical in atomic models of fracture at zero temperature²³.

In order to check whether treating the front and the back of the crack independently at the atomic level led to errors, and to check more generally the validity of our calculations, we have performed a number of full molecular-dynamics calculations in regions of

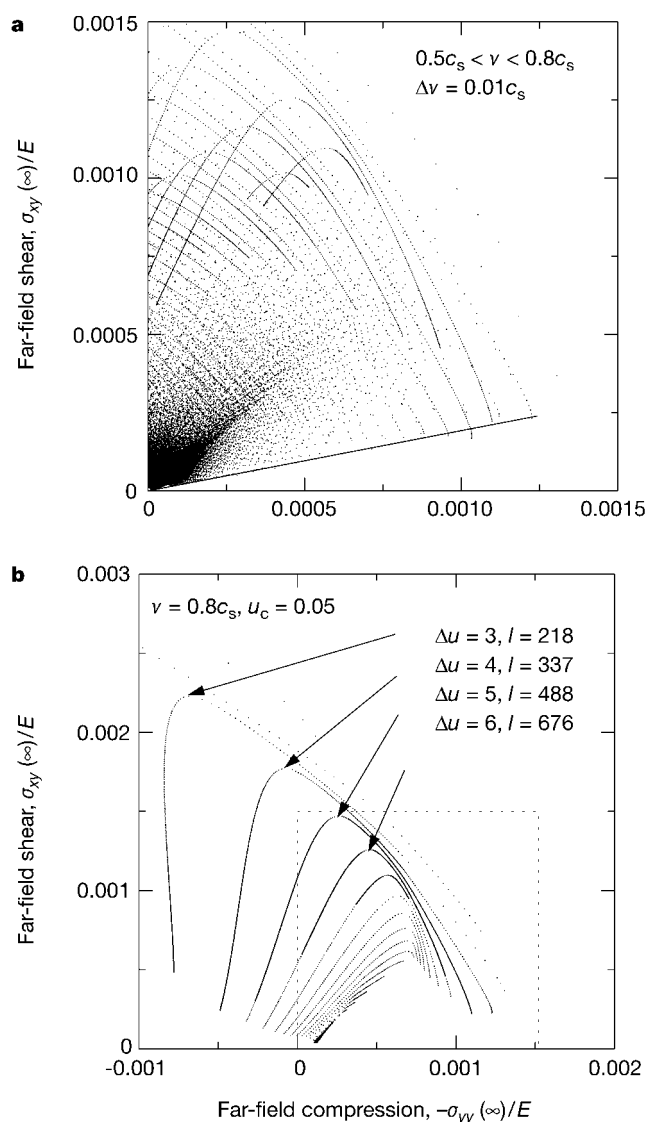


Figure 2 Catalogue of shear $\sigma_{xy}(\infty)$ and compressive $-\sigma_{yy}(\infty)$ stresses, applied to infinitely large systems, that support steadily moving self-healing cracks. **a**, Crack moving at 50–80% of the shear wave speed, in velocity increments of 1%. The minimum shear stress that allows cracks to move for a given compressive stress yields an approximately linear relationship (Coulomb's law), indicated by the straight line in **a**. The slope of this line is the friction coefficient, 0.2. Stresses are normalized by Young's modulus E and

velocities v are normalized by shear wave speed c_s . Integral values of crack length l and total slip Δu lead to a discrete set of states for each velocity v . **b**, As **a** but selecting just the self-healing cracks moving at velocity $v = 0.8c_s$. Points that appear to form a curve have the same slip Δu but different crack lengths l . The region inside the box with dotted outline is shown in **a**.

parameter space where self-healing cracks were predicted to exist. Because of the long running times and complicated initial conditions needed to produce steady states, we have limited ourselves to small systems of about 500,000 atoms. We have found three self-healing cracks consistent with the analytical predictions; a picture of one of them appears in Fig. 1.

With the catalogue of self-healing cracks in hand, we now return to the matter of how they lead to friction. For any given compressive stress ($-\sigma_{yy}(\infty)$) there is a minimum shear stress $\sigma_{xy}(\infty) \approx -0.2\sigma_{yy}(\infty)$ that allows the cracks to begin moving (see Fig. 2a). For any smaller shear stress, the upper block cannot slide; for any larger shear stress, there is a way for it to slide. This sliding is well described by a coefficient of friction to the extent that the lower boundary of states can be approximated by a straight line. The lower boundary of states depends only weakly upon the model parameters, which include bond stiffness and interface strength.

What we have in fact determined is a lower bound on a coefficient of static friction at zero temperature. Static friction will be determined by when self-healing cracks actually initiate, which is not necessarily identical with when they first become possible. Once sliding begins, properties of kinetic friction will be determined by populations of self-healing cracks. The speed at which the upper block slides will depend upon the number of self-healing cracks moving at any time, and is not directly determined by the speeds of these cracks.

The question of when self-healing cracks actually underlie frictional sliding will have to be settled by experiments. There is some evidence for such cracks from experiments aimed at settling questions about earthquakes⁵⁻⁷. Real surfaces are certainly not flat enough over macroscopic lengths for our description of friction to be complete, and other mechanisms of frictional sliding^{1,2,4} will certainly compete with this one. However, we believe that our calculation of a friction coefficient from the atomic scale up constitutes progress. We hope that these definite mathematical results for an ideal case will be useful for proceeding to more realistic ones, and will motivate new experiments. □

Received 7 March; accepted 11 August 2001.

- Persson, B. N. J. *Sliding Friction: Physical Principles and Applications* (Springer, Heidelberg, 1998).
- Comninou, M. & Dundurs, J. Can solids slide without slipping? *Int. J. Solids Struct.* **14**, 251–260 (1978).
- Weertman, J. Unstable slippage across a fault that separates elastic media of different elastic constants. *J. Geophys. Res.* **B 85**, 1455–1461 (1980).
- Caroli, C. Slip pulses at a sheared frictional viscoelastic/nondeformable interface. *Phys. Rev. E* **62**, 1729–1735 (2000).
- Brune, J. N., Brown, S. & Johnson, P. A. Rupture mechanism and interface separation in foam rubber models of earthquakes: a possible solution to the heat flow paradox and the paradox of large overthrust. *Tectonophysics* **218**, 59–67 (1993).
- Brown, S. R. Frictional heating on faults: Stable sliding versus stick slip. *J. Geophys. Res.* **103**, 7413–7420 (1998).
- Anooshehpour, A. & Brune, J. N. Wrinkle-like Weertman pulse at the interface between two blocks of foam rubber with different velocities. *Geophys. Res. Lett.* **26**, 2025–2028 (1999).
- Scott, R. S. Seismicity and stress rotation in a granular model of the brittle crust. *Nature* **381**, 5922–595 (1996).
- Ben-Zion, Y. & Andrews, D. J. Properties and implications of dynamic rupture along a material interface. *Bull. Seismol. Soc. Am.* **88**, 1085–1094 (1998).
- Mora, P. The weakness of earthquake faults. *Geophys. Res. Lett.* **26**, 123–126 (1999).
- Place, D. & Mora, P. The lattice solid model to simulate the physics of rocks and earthquakes: incorporation of friction. *J. Comput. Phys.* **150**, 332–372 (1999).
- Freund, L. B. *Dynamic Fracture Mechanics* (Cambridge Univ. Press, Cambridge, 1990).
- Dhaliwal, R. J. & Saxena, H. S. Moving Griffith crack at the interface of two orthotropic elastic layers. *J. Math. Phys. Sci.* **26**, 237–254 (1992).
- Das, S. & Patra, B. Moving Griffith crack at the interface of two dissimilar orthotropic half planes. *Eng. Fract. Mech.* **54**, 523–531 (1996).
- Ranjith, K. & Rice, J. R. Slip dynamics at an interface between dissimilar materials. *J. Mech. Phys. Solids* **49**, 341–361 (2001).
- Cochard, A. & Rice, J. R. Fault rupture between dissimilar materials: III—posedness, regularization, and slip-pulse response. *J. Geophys. Res.* **B 105**, 25891–25907 (2001).
- Slepyan, L. I. Plane problem of a crack in a lattice. *Izv. Akad. Nauk SSSR Mekh. Tverd. Tela* **16**, 101–115 (1982).
- Marder, M. & Gross, S. Origin of crack tip instabilities. *J. Mech. Phys. Solids* **43**, 1–48 (1995).
- Wilson, G. T. The factorization of matrixial spectral densities. *SIAM J. Appl. Math.* **23**, 420–426 (1972).
- Eggermont, P. & Lubich, C. Fast numerical solution of singular integral equations. *J. Integral Equations Appl.* **6**, 335–351 (1994).

- Williams, M. L. The stresses around a fault or crack in dissimilar media. *Bull. Seismol. Soc. Am.* **49**, 199–204 (1959).
- Deng, X. Complete complex series expansions of near-tip fields for steadily growing interface cracks in dissimilar isotropic materials. *Eng. Fract. Mech.* **42**, 237–242 (1992).
- Holland, D. & Marder, M. Cracks and atoms. *Adv. Mater.* **11**, 793–806 (1999).
- Müser, M. H., Wenning, L. & Robbins, M. O. Simple microscopic theory of Amontons's laws for static friction. *Phys. Rev. Lett.* **86**, 1295–1298 (2001).
- Heaton, T. H. Evidence for and implications of self-healing pulses of slip in earthquake rupture. *Phys. Earth Planet. Interiors* **64**, 1–20 (1990).
- Adams, G. G. Dynamic motion of two elastic half-spaces in relative sliding without slipping. *J. Tribol.* **121**, 455–461 (1999).
- Ranjith, K. & Rice, J. R. Slip dynamics at an interface between dissimilar materials. *J. Mech. Phys. Solids* **49**, 341–361 (2001).

Acknowledgements

We thank H. Swinney for suggestions on presentation. This work was supported by the NSF and by a fellowship from TICAM at The University of Texas at Austin.

Correspondence and requests for materials should be addressed to M.M. (e-mail: marder@chaos.ph.utexas.edu).

A high-strain-rate superplastic ceramic

B.-N. Kim, K. Hiraga, K. Morita & Y. Sakka

National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan

High-strain-rate superplasticity describes the ability of a material to sustain large plastic deformation in tension at high strain rates of the order of 10^{-2} to 10^{-1} s^{-1} and is of great technological interest for the shape-forming of engineering materials. High-strain-rate superplasticity has been observed in aluminium-based¹ and magnesium-based² alloys. But for ceramic materials, superplastic deformation has been restricted to low strain rates of the order of 10^{-5} to 10^{-4} s^{-1} for most oxides^{3,4} and nitrides⁵ with the presence of intergranular cavities leading to premature failure. Here we show that a composite ceramic material consisting of tetragonal zirconium oxide, magnesium aluminate spinel and α -alumina phases exhibits superplasticity at strain rates up to 1 s^{-1} . The composite also exhibits a large tensile elongation, exceeding 1,050 per cent for a strain rate of 0.4 s^{-1} . The tensile flow behaviour and deformed microstructure of the material indicate that superplasticity is due to a combination of limited grain growth in the constitutive phases and the intervention of dislocation-induced plasticity in the zirconium oxide phase. We suggest that the present results hold promise for the application of shape-forming technologies to ceramic materials.

In superplastic materials, the primary deformation mechanism is grain-boundary sliding, and it is the rate of this process that determines the macroscopic strain rate. Because cavitation due to grain-boundary sliding must be accommodated by diffusion and/or dislocation processes for successive deformation, a short accommodation length—which means a small grain size—is indispensable for attaining high-strain-rate superplasticity. For the same reason, stability of the small grain size is also essential. If grain growth occurs actively during deformation, the accommodation length increases and retards facile grain-boundary sliding. This causes an increase in the level of stress necessary for successive deformation—that is, strain-hardening. Strain-hardening enhances the extent of stress concentration on the sliding grain boundaries or grain corners, resulting in the formation of intergranular cavities that leads to premature failure. We have fabricated a multi-phase

ceramic composite consisting of 40 vol.% ZrO_2 , 30 vol.% spinel and 30 vol.% Al_2O_3 ; this material has submicrometre-sized grains and grain growth is strongly suppressed by dispersed phases.

The starting materials were commercial powders of high purity $\alpha\text{-Al}_2\text{O}_3$ (>99.99%, TM-DAR, Taimei Chemical), tetragonal ZrO_2 stabilized by 3 mol% Y_2O_3 (>99.97%, TZ-3Y, Tosoh) and MgO (>99.97%, 100A, Ube Chemical) with nominal particle diameters of 0.2, 0.07 and 0.017 μm , respectively. The powders were mixed in a ball-mill, using balls of pure Al_2O_3 (>99.9%) and ethanol. After drying and granulation with a 60-mesh sieve, the mixed powder was pressed at 40 MPa and then cold-isostatically pressed at 200 MPa. The compacts were sintered at 1,400 $^\circ\text{C}$ for 1 h in air, during which a spinel phase was formed by chemical reaction between the MgO and Al_2O_3 powders. The density of the sintered body was 4.57 g cm^{-3} . As shown in Fig. 1, the sintered microstructure consisted of equiaxed $\alpha\text{-Al}_2\text{O}_3$, spinel and tetragonal ZrO_2 grains. The average grain size, defined as the equivalent area diameter of the grain measured by scanning electron micrography (SEM), was 0.29 μm for the Al_2O_3 phase, and 0.18 μm for both ZrO_2 and spinel phases. The mean grain size of these phases was 0.21 μm .

From the sintered compacts, flat tensile specimens were machined with a gauge length of 8 mm, a width of 3 mm and a thickness of 2 mm. Constant-displacement-rate tensile tests were performed with an Instron-type testing machine at 1,650 $^\circ\text{C}$ in vacuum. Tensile strain and stress were determined from the cross-head displacement by considering the compliance of the testing machine and by assuming uniform elongation in the gauge portion. At the test temperature, the grain sizes just before deformation were about twice the as-sintered ones.

As shown in Fig. 2, tensile deformation in the present material proceeded uniformly without local necking, and tensile elongation reached 390% at an initial strain rate of 1 s^{-1} . To our knowledge, this is the highest strain rate reported for any superplastic ceramic material. At a slightly lower strain rate of 0.4 s^{-1} , the material did not fail even after the tensile elongation reached 1,050%, which is the maximum elongation that can be examined with the present testing machine. An elongation to failure of 1,040%, which is (to our knowledge) the maximum value so far recorded in ceramics, has been attained in a Y_2O_3 -stabilized tetragonal ZrO_2 doped with 5 wt% ($\sim 13 \text{ vol.}\%$) SiO_2 at an initial strain rate of $1.3 \times 10^{-4} \text{ s}^{-1}$

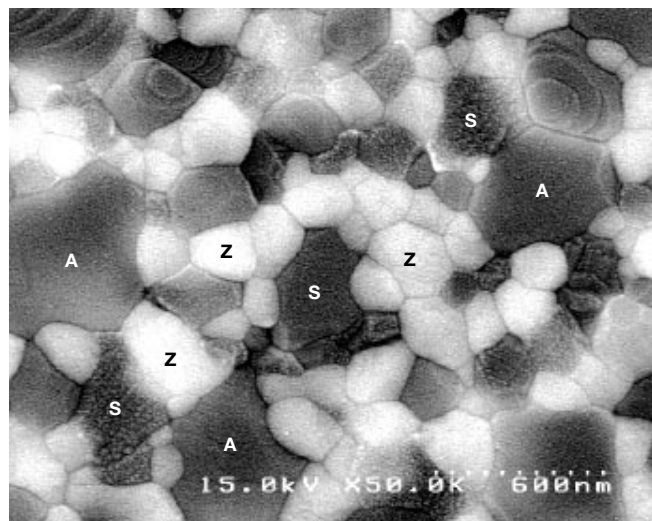


Figure 1 Scanning electron micrograph of the as-sintered microstructure. Grain labels A, S and Z represent Al_2O_3 , spinel and ZrO_2 , respectively. The constitutive phases were identified by X-ray diffraction. The molar Al_2O_3 to MgO ratio in the spinel phase was also determined by X-ray diffraction as 1.1 for the as-sintered material and 2.2 for a specimen quenched from the deformation temperature.

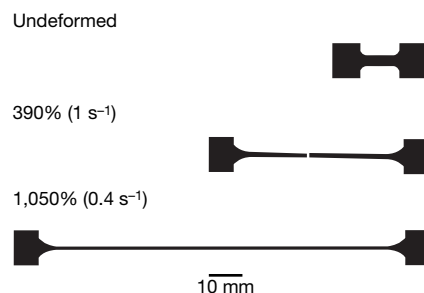


Figure 2 Specimens before and after deformation at 1,650 $^\circ\text{C}$. At 0.4 s^{-1} , the tensile test was terminated before failure owing to a limitation of the loading span available in the testing machine.

(ref. 4). It should be noted, however, that the strain rate to attain the 1,050% elongation in the present material is more than 3,000 times higher than that available for the SiO_2 -doped material. We also note that such a difference in the available strain rate leads to quite a large difference in the deformation time: to attain an elongation of 1,000%, for example, it takes 20 h at $1.3 \times 10^{-4} \text{ s}^{-1}$ in the SiO_2 -doped ZrO_2 , whereas it takes only 25 s at 0.4 s^{-1} in the present material.

Tensile flow data (Fig. 3) exhibit some characteristics different from those of other superplastic ceramics^{1,3-5}. The first point is the lack of strain-hardening in the present material. Flow stress either starts to decrease immediately after yielding (1 s^{-1}), or it remains almost constant up to a tensile strain of 0.8 and then gradually decreases (0.4 s^{-1}). This is in contrast to the general behaviour of other superplastic ceramics, where strain-hardening occurs through grain growth, and is followed by a decrease in flow stress; this decrease is mainly due to a decrease in strain rate with increasing strain under constant-displacement-rate loading, and partially to cavitation at large strains. Microstructural examination of the deformed specimens revealed that the lack of strain-hardening is due mainly to highly limited grain growth. The increment of grain size in each phase was about 10% or less after deformation to failure.

There are two possible explanations for the limited grain growth in the present material; the heightened strain rates and the particle dispersion. The extent of grain growth is proportional to the deformation time and strain⁶. Hence, the shorter the deformation time or the smaller the tensile strain, the smaller the increment of

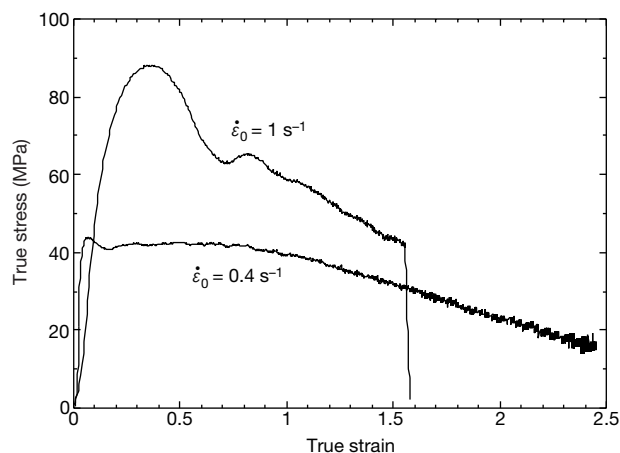


Figure 3 Stress-strain curves at 1,650 $^\circ\text{C}$. $\dot{\epsilon}_0$ is the initial strain rate. We note that the undulated feature after yielding, particularly at $\dot{\epsilon}_0 = 1 \text{ s}^{-1}$, is typical of dynamic recrystallization in metallic materials. The extent of yield-drop increases with increasing strain rate.

grain size—and in our material there is a shortened deformation time due to the heightened strain rates. As far as particle dispersion is concerned, the present material has characteristics different from those of conventional materials: a tri-phase structure, where the amount of each phase is similar. For a given volume of dispersed particles, such a tri-phase structure decreases the frequency of the grain boundaries between the same phases or increases the separation distance between the phases. This situation is of benefit to the suppression of grain growth, as grain growth occurs by the migration of such grain boundaries and/or by the coarsening of grains through interphase boundary diffusion. For the same reason, an increase in the amount of dispersed particles is known to enhance the suppression of grain growth⁷. Indeed, the mean grain size of 0.21 μm in the present 30 vol.% Al_2O_3 –40 vol.% ZrO_2 –30 vol.% spinel is much smaller than that of 80 vol.% Al_2O_3 –10 vol.% ZrO_2 –10 vol.% spinel⁸, where the mean grain size reached 0.45 μm for a similar sintering procedure. Thus, the limited grain growth in the present material can be attributed to both the heightened strain rates and the multi-phase structure.

Another point to note in the data shown in Fig. 3 is the yield-drop appearing just after initial yielding, accompanied by an indication of undulated flow. The yield-drop becomes more prominent with increasing strain rate. Such flow behaviour is known to appear during high-temperature plastic deformation in metallic polycrystals and single crystals owing to dislocation processes including dynamic recrystallization⁹, but not in superplastic (or superplastic-like) polycrystalline ceramics. This non-appearance can be rationalized by the generally accepted considerations that grains in superplastic ceramics behave as rigid bodies and deformation occurs through grain-boundary sliding accommodated by diffusion^{1,3–5}. Thus, the flow behaviour shown in Fig. 3 suggests that other deformation mechanisms may intervene in the high-strain-rate deformation of the present material.

Figure 4 shows a typical example of substructures developed in ZrO_2 grains during high-strain-rate deformation. Dense intergranular dislocations, some of which are tangled, and the occurrence of subgrain boundaries are characteristics of the ZrO_2 grains after deformation. This observation indicates that dislocations were generated and moved in the ZrO_2 grains during deformation. In other words, the ZrO_2 grains are not rigid, but are deformed plastically to some extent during the high-strain-rate deformation.

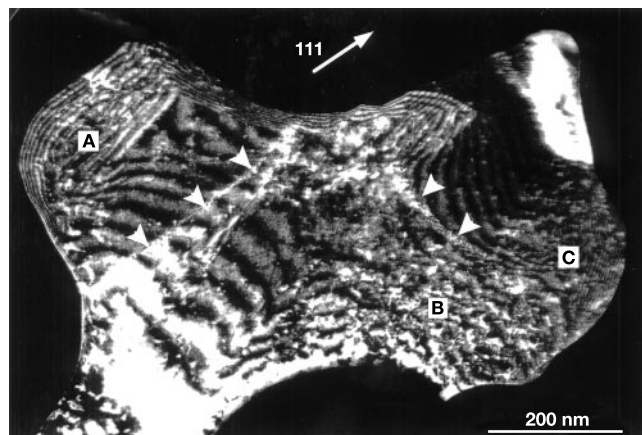


Figure 4 Transmission electron micrograph of a ZrO_2 grain in a specimen deformed to an elongation of 390% at 1,650 °C and at an initial strain rate of 1 s^{-1} . Dislocation substructures developed during high-strain-rate deformation can be seen. Intergranular dislocations, some of which are tangled, are seen around positions A, B and C. White arrow-heads indicate subgrain boundaries. These substructures do not appear in undeformed specimens.

Such dislocation activity has also been found recently in some fine-grained oxides such as spinel¹⁰ and tetragonal ZrO_2 (ref. 11). The observation also indicates that the dislocations were rearranged to form subgrain boundaries during deformation—that is, dynamic recovery occurred in the ZrO_2 grains. The observed features imply that the ZrO_2 grains may contribute to the accommodation process of grain-boundary sliding through their plasticity. This means equivalently that the stress concentration developed during deformation can be relaxed not only by diffusion processes, but also by the plasticity of the ZrO_2 grains. Such a consideration is supported by the following cavitation behaviour in this material.

Cavitation damage in the present material was found to be quite limited during high-strain-rate tensile deformation. Although cavity stringers along the stress axis occurred at large strains, microcracks along the direction normal to the stress axis, which are often found in existing superplastic ceramics¹², were not observed. The average cavity volume fraction measured along the gauge portion was 0.051 and 0.061 for the specimens deformed at 1 s^{-1} (failed at 390% elongation) and 0.4 s^{-1} (elongated to 1,050%), respectively. At 0.1 s^{-1} , the cavity volume fraction was less than 0.025 at an elongation of about 1,000%. These values are quite small compared to those of other superplastic ceramics: in a Y_2O_3 -stabilized tetragonal ZrO_2 , which is a representative superplastic ceramic material, the cavity volume fraction reached 0.3 at an elongation of 700% at 1,550 °C and at a strain rate of $8.3 \times 10^{-5} \text{ s}^{-1}$ (ref. 12). At smaller tensile strains of around 100%, where most actual plastic forming has been performed in metallic materials, the level of cavitation damage can be further suppressed in this material. At a strain rate of 0.4 s^{-1} , for example, the cavity volume fraction after 100% elongation was evaluated to be less than 0.005 from the interpolation of the obtained data, and this value is similar to that which occurs in metal forming¹³.

The suppression of cavity formation, which enables the large elongation in the present material, indicates that stress concentrations generated at grain boundaries and corners are sufficiently relaxed during deformation—if the stress relaxation was insufficient, cavitation would be much enhanced. In conventional superplastic ceramics, the relaxation occurs through diffusion¹⁴, and we expect that this is also the case in the present material. However, the differences in cavitation behaviour and available strain rates between the conventional and the present materials indicate that an additional mechanism is at work in the latter. We attribute the occurrence of the additional relaxation to the above-mentioned dislocation-induced plasticity of the ZrO_2 dispersed among the other constitutive grains.

We thus conclude that the attainment of high-strain-rate superplasticity and the large tensile elongation in the present ZrO_2 –spinel– Al_2O_3 composite arises from the multi-phase structure. This structure enables the achievement of submicrometre-sized grain structure after sintering and strongly suppressed grain growth during deformation. Furthermore, the dislocation-induced plasticity in the ZrO_2 grains is likely to play an additional role in the accommodation of grain-boundary sliding, that is, stress relaxation during high-strain-rate deformation.

The present results show that it is possible to overcome the major drawbacks of conventional superplastic ceramics in engineering applications, namely the low strain rate available for forming¹⁵ and the extensive cavitation. We accordingly consider that, although problems remain to be solved in the area of further suppressing cavitation damage during deformation, the present results provide a way of applying superplasticity to the shape-forming of ceramic materials. □

Received 31 January; accepted 18 July 2001.

- Nieh, T. G., Gilman, P. S. & Wadsworth, J. Extended ductility at high strain rates in a mechanically alloyed aluminum alloy. *Scripta Metall.* **19**, 1375–1378 (1985).
- Mabuchi, M. & Higashi, K. High-strain-rate superplasticity in magnesium matrix composites containing Mg₂Si particles. *Phil. Mag. A* **74**, 887–905 (1996).

3. Wakai, F., Sakaguchi, S. & Matsuno, Y. Superplasticity of yttria-stabilized tetragonal ZrO₂ polycrystals. *Adv. Ceram. Mater.* **1**, 259–263 (1986).
4. Kajihara, K., Yoshizawa, Y. & Sakuma, T. The enhancement of superplastic flow in tetragonal zirconia polycrystals with SiO₂-doping. *Acta Metall. Mater.* **43**, 1235–1242 (1995).
5. Wakai, F. et al. A superplastic covalent crystal composite. *Nature* **344**, 421–423 (1990).
6. Kim, B.-N., Hiraga, K., Sakka, Y. & Ahn, B.-W. A grain boundary diffusion model of dynamic grain growth during superplastic deformation. *Acta Mater.* **47**, 3433–3439 (1999).
7. Smith, C. S. Grains, phases and interfaces: an interpretation of microstructure. *Trans. Metall. Soc. AIME* **175**, 15–51 (1948).
8. Kim, B.-N., Hiraga, K., Morita, K. & Sakka, Y. Superplasticity in alumina enhanced by co-dispersion of 10% zirconia and 10% spinel particles. *Acta Mater.* **49**, 887–895 (2001).
9. McQueen, H. J. & Jonas, J. J. in *Treatise on Materials Science and Technology* Vol. 6 (ed. Arsenault, R. J.) 393–493 (Academic, New York, 1975).
10. Ting, C.-J. & Lu, H.-Y. Hot-pressing of magnesium aluminate spinel—II. microstructure development. *Acta Mater.* **47**, 831–840 (1999).
11. Morita, K. & Hiraga, K. Deformed substructures in fine-grained tetragonal zirconia. *Phil. Mag. Lett.* **81**, 311–319 (2001).
12. Schissler, D. J., Chokshi, A. H., Nieh, T. G. & Wadsworth, J. Microstructural aspects of superplastic tensile deformation and cavitation failure in a fine-grained yttria stabilized tetragonal zirconia. *Acta Metall. Mater.* **39**, 3227–3236 (1991).
13. Pilling, J. & Ridley, N. Effect of hydrostatic pressure on cavitation in superplastic aluminium alloys. *Acta Mater.* **34**, 669–679 (1986).
14. Evans, A. G., Rice, J. R. & Hirth, J. P. Suppression of cavity formation in ceramics: prospects for superplasticity. *J. Am. Ceram. Soc.* **63**, 368–375 (1980).
15. Wittenauer, J. Applications of ceramic superplasticity: challenges and opportunities. *Mater. Sci. Forum* **243–245**, 653–662 (1997).

Acknowledgements

We thank N. Sekine for experimental assistance.

Correspondence and requests for materials should be addressed to B.N.K. (e-mail: KIM.Byung-Nam@nims.go.jp).

Ultrafast holographic nanopatterning of biocatalytically formed silica

Lawrence L. Brott*, Rajesh R. Naik*, David J. Pikas*, Sean M. Kirkpatrick*, David W. Tomlin*, Patrick W. Whitlock*†, Stephen J. Clarson† & Morley O. Stone*

* Materials and Manufacturing Directorate, Air Force Research Laboratory, 3005 P Street, Wright-Patterson Air Force Base, Ohio 45433-7702, USA

† Department of Materials Science and Engineering, University of Cincinnati, 497 Rhodes Hall, Cincinnati, Ohio 45221-0012, USA

Diatoms are of interest to the materials research community because of their ability to create highly complex and intricate silica structures under physiological conditions: what these single-cell organisms accomplish so elegantly in nature requires extreme laboratory conditions to duplicate^{1,2}—this is true for even the simplest of structures. Following the identification of polycationic peptides from the diatom *Cylindrotheca fusiformis*, simple silica nanospheres can now be synthesized *in vitro* from silanes at nearly neutral pH and at ambient temperatures and pressures^{3,4}. Here we describe a method for creating a hybrid organic/inorganic ordered nanostructure of silica spheres through the incorporation of a polycationic peptide (derived from the *C. fusiformis* silaffin-1 protein) into a polymer hologram created by two-photon-induced photopolymerization. When these peptide nanopatterned holographic structures are exposed to a silicic acid, an ordered array of silica nanospheres is deposited onto the clear polymer substrate. These structures exhibit a nearly fifty-fold increase in diffraction efficiency over a comparable polymer hologram without silica. This approach, combining the ease of processability of an organic polymer with the improved mechanical and optical properties of an inorganic material, could be of practical use for the fabrication of photonic devices.

We have recently developed a holographic two-photon-induced photopolymerization (H-TPIP) process⁵ and here we describe how this technique can be used to prepare nanopatterned structures that contain biological macromolecules. Unlike conventional holograms formed through the use of ultraviolet lasers, holograms created through the two-photon process use an ultrafast infrared laser. Because infrared wavelengths typically do not alter the functionality of biological compounds, monomer formulations containing peptides can be polymerized without affecting the biological activity. We incorporated a peptide that has recently been shown to be responsible for biosilification into a formulation to be cured by a holographic two-photon-induced photopolymerization with the expectation that the peptide would be segregated into regions of low crosslinking density. The approach of using ultraviolet lasers to phase separate small liquid crystal molecules in a polymer-based hologram has been used extensively⁶ and we hypothesized that this technique would also be applicable to the H-TPIP process. We predicted that exposing the peptide-containing structure to a liquid silane would cause silica to form in the holographic nanopattern (see Fig. 1) and that this hybrid organic/inorganic device would have a higher degree of order leading to a superior device compared to randomly ordered monolayers of silica on indium-tin oxide (ITO) coated glass⁷.

A short 19-amino-acid R5 peptide unit (SSKKSGSYSGKSGSKRR IL) of the silaffin-1 precursor polypeptide from *C. fusiformis* is able to catalyse the formation of silica nanospheres within minutes when added to silicic acid at neutral pH and ambient temperature³. A chemically synthesized R5 peptide that lacks a post-translational modification of its lysine residues was used in the present work. The post-translational modification of lysine residues is required for silica formation under acidic pH conditions^{3,8}. However, because our research was conducted under slightly basic conditions, the modification of the lysine residues was unnecessary. Consequently, work began by incorporating this peptide (0.80 mg in 16 µl of water) into a monomer formulation. This formulation consisted of 160 µl SR-9035, 0.022 g SR-399 (SR-9035 is a trimethylolpropane triacrylate and SR-399 is a dipentaerythritol pentaacrylate obtained from Sartomer which were used without the removal of inhibitor), 0.006 g triethanol amine and 0.005 g isopropyl thioxanthone; the entire mixture was heated for 15 min at 50 °C to aid in dissolution. The triacrylate was chosen for its high water miscibility which is due to its numerous ethylene glycol units, and the pentaacrylate was used to create a highly crosslinked system. The triethanol amine functions as a coinitiator and thioxanthone as the initiator. Typically, in a two-photon-initiated polymerization, a fluorescent chromophore is also required to absorb two photons of near-infrared laser light. The excited chromophore transfers its energy to the initiator which begins the polymerization process. However, we have found that the thioxanthone used in this formulation does not require highly coloured chromophores, and consequently, extremely large curing depths and exceptionally clear and colourless polymers are produced^{9,10}.

A thin layer (175 µm) of the monomer/peptide formulation was deposited onto a clean glass slide which was then placed in a miniature atmospheric chamber fitted with glass windows and

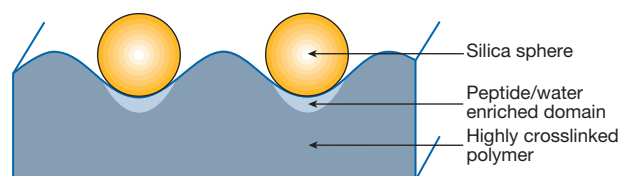


Figure 1 Cross-section of the hologram. The peptide-rich regions that are formed during the holographic polymerization process are shown.

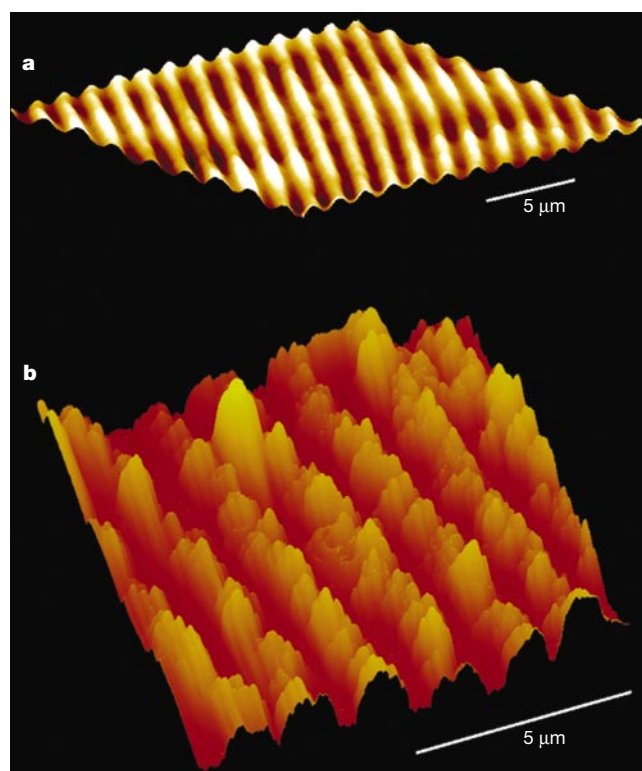


Figure 2 Surface relief pattern of the cured polymer. Atomic force microscope images of the control polymer before silification (**a**) and the hybrid structure after silification (**b**).

flushed with nitrogen. The sample was cured in a two-beam transmission holographic arrangement using a 790-nm titanium-sapphire laser (90-fs pulse width with a repetition rate of 500 Hz) for 30 s. The intensity distribution of the volume hologram drives the local polymerization rate as a function of the local field intensity, which results in alternating areas of high and low crosslink density. Because certain areas of the sample cure more rapidly than others, the smaller molecules (namely water and peptide) phase separate from the areas of higher crosslink density and migrate into areas of lower density. This phenomenon has been observed in similar systems using liquid crystals as the small molecule⁶. An alternative explanation of this phase separation could be that as the hydrophilic monomer is converted into a more hydrophobic polymer, the peptide is driven into the monomer-rich regions. As a result, peptide-rich domains are created in the polymer sample with the periodicity of the hologram. After the curing process, the sample was briefly rinsed with water to remove any uncured monomer. Atomic force microscopy (AFM) revealed that the hologram had a periodicity of 1.33 μm (see Fig. 2a).

The silane precursor (1 M tetrahydroxysilane) was synthesized by dissolving tetramethyl orthosilicate (TMOS) in 1 mM HCl. This product was then added to a sodium phosphate–citrate buffer (pH 8) to produce a final concentration of 113 mM. We note that this dilute solution remains stable for over two hours, after which it slowly converts into a clear amorphous gel. Freshly prepared hydrolysed silane was slowly applied to the hologram and allowed to react with the R5 peptide embedded in the hologram for 10 min before being rinsed with water to remove any unreacted silane. A control hologram lacking the R5 peptide was also treated with the tetrahydroxysilane solution but did not exhibit any nanosphere formation (see Fig. 3a). However, when a sample that included the peptide and was treated with the silane was analysed by scanning electron microscope, it was revealed that silica spheres formed a regular two-dimensional array with the periodicity of the hologram (see Fig. 3b). A study of the size distribution of the silica spheres

reveals that the average nanosphere diameter is 452 nm (± 81 nm). The silica content of the spheres was confirmed using electron dispersive spectroscopy (EDS). Additionally, analysis using the AFM indicated that the hologram had a periodicity of 1.60 μm with the silica spheres embedded in the troughs of the surface relief pattern (see Fig. 2b). The difference in the spacing between the holograms treated with and without the tetrahydroxysilane solution can be explained by the fact that the control grating shrinks as it dries out owing to water evaporation, whereas the shrinkage in the hybrid hologram is inhibited owing to the added mechanical strength of the silica spheres, preventing the ridges of the hologram from moving closer together. Consequently, the untreated grating exhibited nearly 17% more shrinkage than the treated grating. Also, the silica spheres are the most prominent feature of the hologram and the troughs in the structure are actually the peaks of the polymer.

Finally, to test the improvement that this technique can impart to an optical device, the first-order diffraction efficiency of the treated hologram was compared to that of the untreated sample. These measurements were performed by transmitting a helium-neon laser through each sample and measuring the diffraction pattern in the far field. A measurement of the incident and transmitted power in the first-order diffraction spot showed a substantial increase in the diffraction efficiency of the grating with silica versus the grating without, as would be expected from the difference in index and shrinkage. The untreated grating exhibited a diffraction efficiency of approximately 0.02%, while the grating with the silica spheres showed an efficiency of approximately 0.95%. This large increase

Table 1 Properties of holograms with and without silica nanospheres

	Hologram with silica	Hologram without silica
Grating spacing	1.60 μm	1.33 μm
Diffraction efficiency	0.95%	0.019%

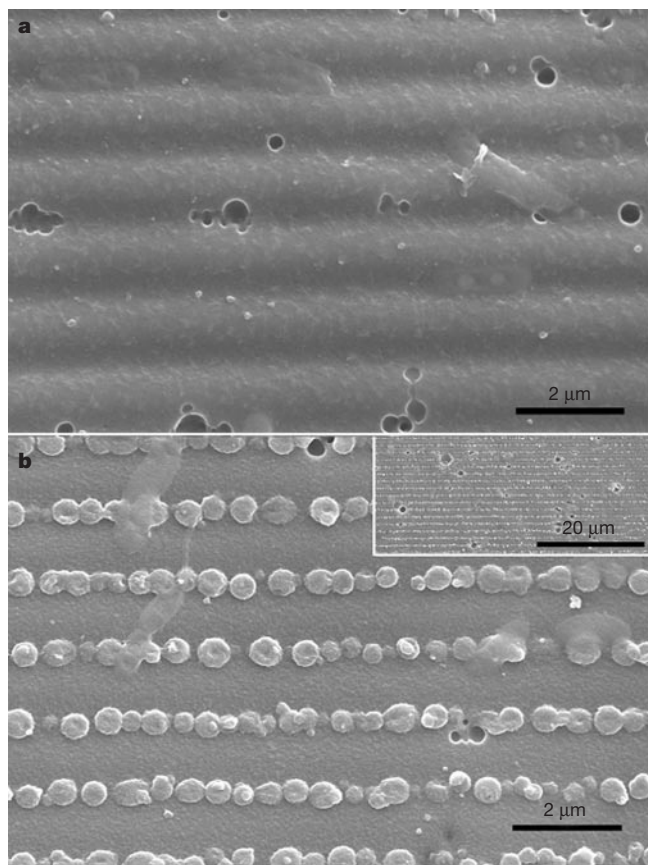


Figure 3 Two-dimensional array of ordered silica nanospheres formed within the hologram. Scanning electron microscope images of the control hologram after being treated with the liquid silane (**a**) and the biosilica nanostructure created by reacting the silane with a peptide-embedded hologram (**b**).

can be attributed to the fact that the spheres form an almost continuous line of silica along the valleys of the hologram, achieving a high fill factor. These data are summarized in Table 1.

We have thus shown that the incorporation of the peptide responsible for biosilification into a microfabricated structure using H-TPIP can result in an unusual composite organic/inorganic device that has significantly improved optical performance and superior mechanical properties compared to those of a corresponding polymeric device without silica. Although we have used a polymer/silica hybrid structure, this technique is universally applicable for any catalyst or binding agent that can be incorporated into a polymer. For example, as different catalysts are identified, a wide variety of unique hybrid structures are now possible with differing shapes and mechanical properties. Additionally, antibodies can be incorporated into the hologram and potentially used to optically identify specific antigens. Consequently, this technique allows a simple yet general and easily modifiable method for nanopatterning. □

Received 18 June; accepted 20 July 2001.

1. Parkinson, J. & Gordon, R. Beyond micromachining: the potential of diatoms. *Trends Biotechnol.* **17**, 190–196 (1999).
2. Morse, D. E. Silicon biotechnology: harnessing biological silica production to construct new materials. *Trends Biotechnol.* **17**, 230–232 (1999).
3. Kröger, N., Deutzmann, R. & Sumper, M. Polycationic peptides from diatom biosilica that direct silica nanosphere formation. *Science* **286**, 1129–1132 (1999).
4. Cha, J. N., Stucky, G. D., Morse, D. E. & Deming, T. J. Biomimetic synthesis of ordered silica structures mediated by block copolypeptides. *Nature* **403**, 289–292 (2000).
5. Kirkpatrick, S. M. *et al.* Holographic recording using two-photon-induced photopolymerization. *Appl. Phys. A* **69**, 461–464 (1999).

6. Bunning, T. J. *et al.* The morphology and performance of holographic transmission gratings recorded in polymer dispersed liquid crystals. *Polymer* **36**, 2699–2708 (1995).
7. Wang, C. *et al.* Two-dimensional ordered arrays of silica nanoparticles. *Chem. Mater.* **12**, 3662–3666 (2000).
8. Kröger, N., Deutzmann, R. & Sumper, M. Silica-precipitating peptides from diatoms, the chemical structure of silaffin-1a from *Cylindrotheca fusiformis*. *J. Biol. Chem.* **276**, 26066–26070 (2001).
9. Belfield, K. D. *et al.* Multiphoton-absorbing organic materials for microfabrication, emerging optical applications and non-destructive three-dimensional imaging. *J. Phys. Org. Chem.* **13**, 837–849 (2000).
10. Brott, L. L., Naik, R. R., Kirkpatrick, S. M., Pikas, D. J. & Stone, M. O. Near-IR two-photon induced polymerizations using either benzophenone or thioxanthone-based photoinitiators. *Polymer Preprints* **42**, 675–676 (2001).

Acknowledgements

This research was supported by the Air Force Office of Scientific Research and by funds to S.J.C. from Dayton Area Graduate Studies Institute to facilitate this UC/AFRL collaboration.

Correspondence should be addressed to M.S.
(e-mail: morley.stone@afri.af.mil).

High-resolution record of climate stability in France during the last interglacial period

Patrick Rioual⁺, Valérie Andrieu-Ponel[†], Miri Rietti-Shati[‡],
Richard W. Battarbee⁺, Jacques-Louis de Beaulieu[†], Rachid Cheddadi[†],
Maurice Reille[†], Helena Svobodova[§] & Aldo Shemesh[‡]

^{*} Environmental Change Research Centre, University College London,
26 Bedford Way, London WC1H 0AP, UK

[†] Institut Méditerranéen d'Ecologie et Paléocologie (UMR-CNRS 6116),
UDSAM, Faculté des Sciences et Techniques de Saint Jérôme,
13397 Marseille Cedex 20, France

[‡] Department of Environmental Sciences and Energy Research,
The Weizmann Institute of Science, Rehovot 76100, Israel

[§] Botanical Institute, CZ-25243 Pruhonice, Prague, Czech Republic

The last interglacial period (127–110 kyr ago) has been considered to be an analogue to the present interglacial period, the Holocene, which may help us to understand present climate evolution. But whereas Holocene climate has been essentially stable in Europe, variability in climate during the last interglacial period has remained unresolved, because climate reconstructions from ice cores^{1,2}, continental records^{3,4} and marine sediment cores^{5,6} give conflicting results for this period⁷. Here we present a high-resolution multi-proxy lacustrine record of climate change during the last interglacial period, based on oxygen isotopes in diatom silica, diatom assemblages and pollen–climate transfer functions from the Ribains maar in France. Contrary to a previous study⁸, our data do not show a cold event interrupting the warm interglacial climate. Instead, we find an early temperature maximum with a transition to a colder climate about halfway through the sequence. The end of the interglacial period is clearly marked by an abrupt change in all proxy records. Our study confirms that in southwestern Europe the last interglacial period was a time of climatic stability and is therefore still likely to represent a useful analogue for the present climate.

In France, the sites La Grande Pile, Les Echets, Lac du Bouchet and Ribains contain continuous sedimentary records from the present to at least the penultimate glacial period (the Saalian glacial). At each site, the last interglacial (the Eemian) can be unequivocally identified^{4,9,10}. At Ribains maar, the age and duration of the Eemian (local name: Ribains interglacial) has been estimated on the basis of correlation with the marine isotope stages (MISs) in

deep-sea cores, and accordingly corresponds to the entire MIS 5e and to the lower part of MIS 5d, 127–110 kyr before present (BP)^{4,11}.

In 1988, a 54-m core was retrieved from Ribains maar. The sediment between depths of 22.50 and 32 m represents a thick, undisturbed and homogeneous diatomite layer, which has been shown (on the basis of pollen analyses) to span the Eemian and early Weichselian glacial¹⁰. The available diatom-rich material provided an opportunity not only to perform ultrahigh-resolution (5-mm interval) diatom analysis across the entire period, but also to produce an independent climate reconstruction using analysis of oxygen isotopes in biogenic silica.

In lacustrine studies, the authigenic origin of diatom silica gives it an important advantage over carbonates¹². Diatom $\delta^{18}\text{O}$, $\delta^{18}\text{O}_{\text{Si}}$, depends on the isotopic composition of the water ($\delta^{18}\text{O}_{\text{w}}$) from which it was secreted and on ambient temperature (with a dependence of $-0.5\text{‰}\text{ }^{\circ}\text{C}^{-1}$)¹³, and has been shown to retain its isotopic signal in the ocean on timescales of at least 430 kyr (ref. 14). The $\delta^{18}\text{O}_{\text{w}}$ of a lake depends on the $\delta^{18}\text{O}$ of the influx, the hydrological setting (mainly the precipitation to evaporation ratio, P/E) and the local climate. Changes in the $\delta^{18}\text{O}$ of the local precipitation replenishing a lake occur because of variations of ambient surface temperatures or modifications in atmospheric circulation. Such modifications may either shift the moisture-source area or change the vapour transport efficiency. The present dependence of the $\delta^{18}\text{O}$ of the precipitation in mid-latitudes on surface temperatures is $0.58\text{‰}\text{ }^{\circ}\text{C}^{-1}$ (ref. 15). A similar dependency has been suggested in

Europe for the period 35–30 kyr BP during the last glacial¹⁶, but for older periods this relationship is unknown. In the sequence we report here, most diatoms are planktonic and can be used qualitatively—through their changes in seasonal production—to indicate changes in the length and intensity of water column stratification and mixing. Pollen analysis, the third proxy used, was carried out at 2.5-cm intervals and provides independent reconstructions of temperature and precipitation¹⁷.

The combined data (Fig. 1) and the palaeoclimate reconstruction (Fig. 2) show that the landscape at the end of the Saalian glacial period was characterized by an open vegetation (low ratio of arboreal to non-arboreal pollen, AP/NAP), low annual precipitation (P_{ann}) and high annual temperature amplitude (see Fig. 2 legend for details). Diatom assemblages were dominated by *Cyclotella rossii*, a species commonly reported in lakes of northern latitudes¹⁸.

At ~ 31.7 m depth in the core, loss on ignition (LOI) and diatom concentration increase simultaneously with abundances of *Aulacoseira subarctica* and then *Stephanodiscus minutulus*. These two planktonic species become dominant, suggesting longer periods of spring mixing¹⁹. Pollen data, however, still indicate late-glacial conditions. Below 31.7 m, samples could not be analysed for $\delta^{18}\text{O}_{\text{Si}}$ owing to the low diatom concentration.

The beginning of the Eemian interglacial, as defined by pollen analysis, occurs at ~ 31.5 m depth in the core. It is marked by high percentages of *Stephanodiscus*, indicating a short period of ice cover

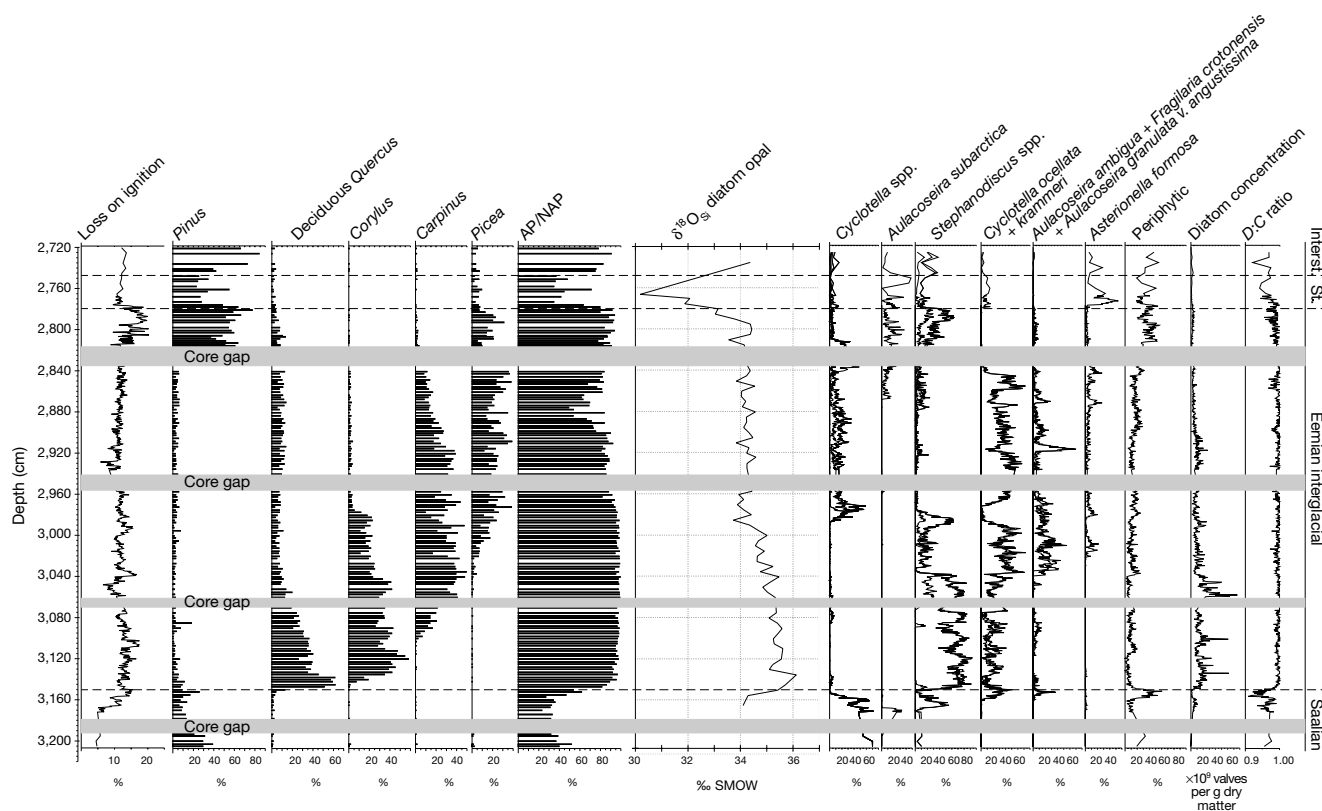


Figure 1 Combined data for the Ribains sediment core plotted against depth. From left to right are shown the percentage of loss on ignition at 500 °C, abundances of selected pollen taxa, the ratio between the sums of arboreal pollen and non-arboreal pollen (AP/NAP) and the stable isotope results of $\delta^{18}\text{O}$ in diatom opal expressed as per mil deviations from the standard mean ocean water composition (‰ SMOW). Further right are shown selected taxa and groups of taxa. *Cyclotella* spp. includes *C. rossii*, *C. radiosa*, *C. bodanica*, *C. distinguenda* var. *unipunctata* and several morphotypes of *C. comensis*. *Stephanodiscus* spp. includes *S. minutulus*, *S. medius* and *S. parvus*. The summer-blooming diatoms *Aulacoseira ambigua*, *Fragilaria crotonensis* and *Aulacoseira granulata*

var. *angustissima* were grouped together as indicators of wind-driven turbulence. The diatom concentration, calculated using microspheres as markers²⁹, is expressed as number of valves per gram of dry matter. The ratio between diatom frustules and chrysophyte cysts, the $D:C$ ratio³⁰, is shown in the rightmost plot. In between each core section of 1 m, there is a gap, 10–20 cm thick, due to loss of material during coring (shaded horizontal bars). The dashed horizontal lines represent the transitions between the different chronostratigraphical zones as defined by pollen analysis shown on the right-hand side (Interst., interstadial St Germain 1; St., stadial Melisey I).

and a low silica-to-phosphorus ratio, and by *Cyclotella ocellata* + *krammeri*, which are characteristic of a strongly stratified water column during late spring and summer¹⁹. This assemblage, along with high diatom concentrations, indicates a productive lake system under a climate characterized by mild winters and calm, hot summers. This is consistent with isotope enrichment from ~34‰ up to ~36‰, at the base of the $\delta^{18}\text{O}_{\text{Si}}$ record, probably due to enrichment of the local precipitation $\delta^{18}\text{O}$. The effect of changing precipitation here overwhelms the temperature dependence of the biogenic silica²⁰. A decrease in *P/E* of the lake is an unlikely explanation, because the pollen reconstruction simultaneously indicates high P_{ann} and a low annual temperature amplitude. This warm event at the beginning of the Eemian, which appears in two other pollen records from the French Massif Central¹⁹, is consistent with the $\delta^{18}\text{O}_{\text{Si}}$ enrichment. Records of $\delta^{18}\text{O}_{\text{carbonate}}$ from central and northern Europe are also characterized by maximum values at the beginning of the Eemian^{12,21}.

At a depth of 31.0 m, the rise in *Carpinus* percentages points to a more continental climate, characterized by lower P_{ann} , lower mean annual temperature (T_{ann}) and greater annual temperature amplitude. At 30.4 m, the expansion of *Picea* is simultaneous with an abrupt decrease of *Stephanodiscus* taxa, an increase in *Cyclotella ocellata* + *krammeri*, a drop in diatom concentration, and increased LOI and percentages of periphytic diatoms. There are also increases in *Fragilaria crotonensis* and *Aulacoseira granulata* var. *angustissima*, two summer/autumn blooming planktonic diatoms that require high silica-to-phosphorus ratio and turbulent conditions to remain in suspension in the water column. This suggests a shorter period of spring mixing and windier summer conditions.

Midway through the Eemian, at 30.0 m depth, pollen analyses indicate a shift towards a colder (especially in summer) and wetter climate, with reduced continentality. Simultaneously and consistently there is a 1‰ drop in $\delta^{18}\text{O}_{\text{Si}}$ and diatom assemblages, characterized by large oscillations between *Stephanodiscus* and

Cyclotella taxa, suggesting abrupt changes in the seasonal mixing regime of the lake. Later, sharp oscillations in reconstructed values for P_{ann} , due to small shifts in the frequencies of *Carpinus* and *Picea*, are probably an artefact of the pollen transfer function. Reconstructed T_{ann} , however, remains stable. The $\delta^{18}\text{O}_{\text{Si}}$ record also remains relatively invariant, suggesting that the changes in the amount of precipitation are balanced by changes in evaporation from the lake, thereby keeping constant the isotopic composition of lake water. The main diatom changes occur at 29.80 m, where the appearance of several *Cyclotella* taxa coincides with the almost complete disappearance of *Corylus* pollen. The mid-Eemian shift towards colder and wetter climate, simultaneously recorded by all three proxies in Ribains, has been identified in other European sequences^{21–23}. This shift observed in continental records can be associated with the closure of the Baltic Sea–White Sea connection due to isostatic uplift²¹ and/or to the decreased sea surface temperature and changing ocean circulation inferred from marine cores from northern Europe²⁴.

The next notable feature occurs between depths of ~29.0 and 28.6 m: an increase in non-arboreal pollen percentages is simultaneous with increased percentages of periphytic diatoms and lower diatom concentration, suggesting lower primary productivity. At the corresponding time period in the Lac du Bouchet sediment sequence, an intra-interglacial cooling has been inferred, mainly from rock magnetism⁸, a feature that was used to support evidence for climate instability in correlation with the oxygen isotope data of a Greenland ice core¹. However, this cooling has not been confirmed by a pollen-based quantitative climate reconstruction⁴ or by subsequent studies of rock-magnetic properties in two other crater lakes in the Massif Central (that is, Saint-Front and Ribains)²⁵. Our multi-proxy data from the Ribains sequence, along with other studies on European lake-sediment records^{12,21,22}, also cast doubt on the climatic significance of this event.

Towards the end of the Eemian (at 27.95 m depth in the core),

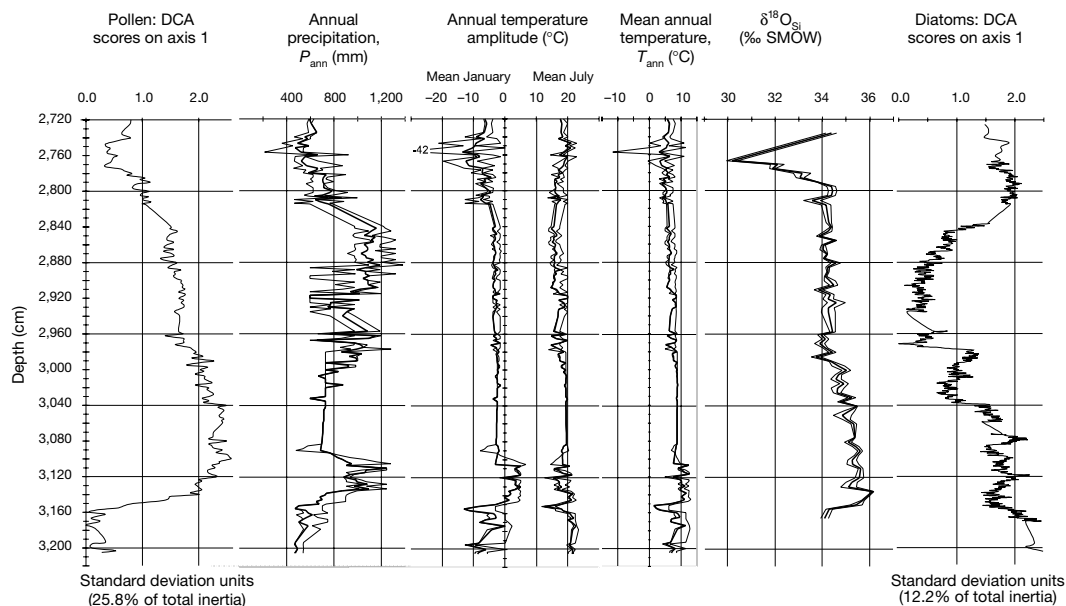


Figure 2 Ribains maar data of detrended correspondence analysis (DCA) for pollen and diatom assemblages and pollen-derived palaeoclimate reconstruction plotted against depth. The plots of DCA scores (on axis 1), expressed as standard deviation of turnover units, provide a summary of the main compositional changes of the diatom and pollen assemblages. Palaeoclimatic pollen reconstructions were obtained using the 'best modern analogue' statistical method, which translates quantitative pollen data into quantitative palaeoclimate estimates¹⁷. The quality of the analogue is measured by the chord distance, and the best ones are selected by Monte Carlo simulation⁴. For all

pollen-based reconstructions, the dotted lines represent the standard error computed by simulation. The annual temperature amplitude, a measure of the continentality of the climate³, was calculated by subtracting mean January temperature (T_{Jan}) from mean July temperature (T_{Jul}). The $\delta^{18}\text{O}_{\text{Si}}$ record, with standard deviation, is also shown for comparison. Note that the reconstructed temperature during the Saalian glacial and the Melisey I stadial is overestimated because the analogue matching routine mistakenly identifies the pollen assemblage in the Saalian as warm temperate grassland rather than cold steppe.

risers in *Picea*, *Stephanodiscus* and LOI indicate a short warming event, although the $\delta^{18}\text{O}_{\text{Si}}$ record, with a sharp decrease from $\sim 34.4\text{‰}$ to 33.1‰ , suggests a cooling. This represents the only major discrepancy among the different proxies at Ribains. However, a short warming trend has been inferred from other European pollen and isotope records^{12,22}.

At Ribains, the transition from the temperate climate of the Eemian to the Melisey I stadial occurred within 900–1,900 years, depending on whether the Eemian lasted 11 or 23 kyr on the continents^{11,23} and assuming a constant sediment accumulation rate in the lake. This contrasts with an estimate²⁶ from La Grande Pile pollen record that this transitional phase lasted less than 400 years.

At 27.8 m depth, the beginning of the last glacial (Melisey I) is characterized by a sharp rise in non-arboreal pollen (including steppe taxa) and a sudden increase in *Asterionella formosa*, a sharp drop in *Stephanodiscus* and LOI, and the re-appearance of summer-blooming *Cyclotella*. *Asterionella formosa* is a species favoured by late-spring overturn after a long period of ice cover²⁷. The constant ratio of euplanktonic to periphytic diatoms and the abundance of *Cyclotella* species, which suggest that the lake level was sufficiently high to allow for summer stratification to take place, indicate that no major changes in lake level²⁸ occurred during this abrupt climate transition. The $\delta^{18}\text{O}_{\text{Si}}$ record shows a further depletion (to about 30.2‰), which could have been caused by a substantial increase in the *P/E* ratio or by a large depletion in the $\delta^{18}\text{O}$ of ground water through the establishment of a colder climate and its cooling effect on precipitation. The pollen reconstruction, which indicates low P_{ann} for this period, points to the second hypothesis. The Ribains depletion at the transition to the Melisey I (4‰) is simultaneous with an abrupt drop in North Atlantic sea surface temperature and the breakdown of the North Atlantic thermohaline circulation²¹. An increase in the temperature difference between the moisture source and the site leads to a decrease in the proportion of the original moisture arriving, and consequently reduces the $\delta^{18}\text{O}$ of precipitation. At ~ 27.45 m, the two uppermost $\delta^{18}\text{O}_{\text{Si}}$ samples at the onset of the Saint Germain I interstadial have enriched $\delta^{18}\text{O}_{\text{Si}}$ values of about 34‰ , similar to the values recorded at the end of the Eemian. Pollen and diatom assemblages also characterize a return to the conditions that prevailed before the Melisey I stadial.

The main climate events of the period investigated (that is, the thermal optimum at the start of the Eemian, mid-Eemian shift, Melisey I) are simultaneously recorded by all three parameters (pollen, diatoms and $\delta^{18}\text{O}_{\text{Si}}$), suggesting that there was no substantial time lag between their responses to climate forcing. From the maximum value measured at the start of the Eemian to the transition with the Melisey I, $\delta^{18}\text{O}_{\text{Si}}$ decreases by about 3‰ . Assuming that the isotopic composition of lake water, controlled by the isotopic composition of the local precipitation, overwhelms the temperature dependency of the silica, this corresponds to a temperature decrease of $\sim 6^\circ\text{C}$. This is consistent with the trend in T_{ann} derived from pollen analysis. The $\delta^{18}\text{O}_{\text{Si}}$ record, which provides an integrated signal of overall climate trends, shows low instability at the scale of the whole interglacial period. It appears, however, less sensitive to climate fluctuations of small amplitude than diatom assemblages or to a certain extent, pollen analysis.

General interglacial stability indicated by this study has also been shown in palaeoceanographic studies^{5,6}, and has been inferred to be a function of switches in state linked to changes in deep-water circulation⁶ and changes in the size and geometry of polar ice sheets⁵.

The comparison of our record with other continental and marine sequences depends on whether the period termed the Eemian in these studies covers the same time interval as our study. It is possible that only the older stage of the temperate period in the Ribains record corresponds to the minimum ice-volume period MIS 5e and to the elapsed part of the Holocene⁷. If this is correct, climate in central and southern Europe was warm and stable during the first 10–12 kyr of the Eemian, and somewhat colder but still stable

during the ~ 10 kyr that followed. This implies that the next shift of the global climate system might be towards a somewhat colder but still stable climate, rather than a shift towards a full glacial. In order to explore the likelihood of such a scenario it is necessary to establish better the chronology of the continental Eemian and its relation to MIS 5.

Note added in proof: Until immediately before publication, we failed to reference, or inform the editors of *Nature*, of an earlier paper of ours³¹ which draws identical conclusions from a subset of the data presented here. □

Received 25 April; accepted 19 July 2001.

- GRIP members. Climate instability during the last interglacial period recorded in the GRIP ice core. *Nature* **364**, 203–207 (1993).
- Grootes, P. M., Stuiver, M., White, J. W. C., Johnsen, S. & Jouzel, J. Comparison of oxygen isotope records from the GISP2 and GRIP ice cores. *Nature* **366**, 552–554 (1993).
- Aalberg, G. & Litt, T. Multiproxy climate reconstructions for the Eemian and Early Weichselian. *J. Quat. Sci.* **13**, 367–390 (1998).
- Cheddadi, R. *et al.* Was the climate of the Eemian stable? A quantitative climate reconstruction from seven European pollen records. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **143**, 73–85 (1998).
- McManus, J. F., Oppo, D. W. & Cullen, J. L. A 0.5-million-year record of millennial-scale climate variability in the North Atlantic. *Science* **283**, 971–975 (1999).
- Cortijo, E., Labeyrie, L., Elliot, M., Balbon, E. & Tisnerat, N. Rapid climatic variability of the North Atlantic Ocean and global climate: a focus of the IMAGES program. *Quat. Sci. Rev.* **19**, 227–241 (2000).
- Kukla, G. J. The last interglacial. *Science* **287**, 987–988 (2000).
- Thouveny, N. *et al.* Climate variations in Europe over past 140 kyr deduced from rock magnetism. *Nature* **371**, 503–506 (1994).
- Reille, M. *et al.* Pollen analytical biostratigraphy of the last five climatic cycles from a long continental sequence from the Velay region (Massif Central, France). *J. Quat. Sci.* **15**, 665–685 (2000).
- de Beaulieu, J. L. & Reille, M. Long Pleistocene pollen sequences from the Velay Plateau (Massif Central, France). I. Ribains maar. *Vegetat. Hist. Archaeobot.* **1**, 233–242 (1992).
- Kukla, G., McManus, J. F., Rousseau, D.-D. & Chuine, I. How long and how stable was the last interglacial? *Quat. Sci. Rev.* **16**, 605–612 (1997).
- Litt, T., Junge, F. W. & Böttger, T. Climate during the Eemian in north-central Europe—a critical review of the palaeobotanical and stable isotope data from central Germany. *Vegetat. Hist. Archaeobot.* **5**, 247–256 (1996).
- Juillet-Leclerc, A. & Labeyrie, L. Temperature dependence of the oxygen isotopic fractionation between diatom silica and water. *Earth Planet. Sci. Lett.* **84**, 69–74 (1987).
- Shemesh, A., Charles, C. D. & Fairbanks, R. G. Oxygen isotopes in biogenic silica: global changes in ocean temperature and isotopic composition. *Science* **256**, 1434–1436 (1992).
- Rozanski, K., Araguás-Araguás, L. & Gonfiantini, R. in *Climate Change in Continental Isotopic Records* (eds Swart, P. K., Lohmann, K. C., McKenzie, J. & Savin, S.) 1–36 (American Geophysical Union, Washington DC, 1993).
- Beyerle, U. *et al.* Climate and groundwater recharge during the last glaciation in an ice-covered region. *Science* **282**, 731–734 (1998).
- Guiot, J., Pons, A., de Beaulieu, J. L. & Reille, M. A 140,000-year continental climate reconstruction from two European pollen records. *Nature* **338**, 309–313 (1989).
- Weckström, J., Korhola, A. & B. Iom, T. The relationship between diatoms and water temperature in thirty subarctic Fennoscandian lakes. *Arct. Alp. Res.* **29**, 75–92 (1997).
- Stoermer, E. F. Evaluating diatom succession: some peculiarities of the Great Lakes case. *J. Paleolimnol.* **8**, 71–83 (1993).
- Rosqvist, G. C., Rietti-Shati, M. & Shemesh, A. Late glacial to middle Holocene climatic record of lacustrine biogenic silica oxygen isotopes from a Southern Ocean island. *Geology* **27**, 967–970 (1999).
- Björck, S. *et al.* Eemian lake development, hydrology and climate: a multi-stratigraphic study of the Hollerup site in Denmark. *Quat. Sci. Rev.* **19**, 509–536 (2000).
- Zagwijn, W. H. An analysis of Eemian climate in western and central Europe. *Quat. Sci. Rev.* **15**, 451–469 (1996).
- Frogley, M. R., Tzedakis, P. C. & Heaton, T. H. E. Climate variability in Northwest Greece during the Last Interglacial. *Science* **285**, 1886–1889 (1999).
- Fronval, T. & Jansen, E. Rapid changes in ocean circulation and heat flux in the Nordic Seas during the last interglacial period. *Nature* **383**, 806–810 (1996).
- Stockhausen, H. & Thouveny, N. Rock-magnetic properties of Eemian maar lake sediments from Massif Central, France: a climatic signature? *Earth Planet. Sci. Lett.* **173**, 299–313 (1999).
- Woillard, G. Abrupt end of the last interglacial s.s. in north-east France. *Nature* **281**, 558–562 (1979).
- Maberly, S. C. *et al.* The rise and fall of *Asterionella formosa* in the South Basin of Windermere: analysis of a 45-year series of data. *Freshwat. Biol.* **31**, 19–34 (1994).
- Wolfin, J. A. Late Holocene lake-level and lake development signals in Lower Herring Lake, Michigan. *J. Paleolimnol.* **15**, 19–45 (1996).
- Battarbee, R. W. & Kneen, M. J. The use of electronically counted microspheres in absolute diatom analysis. *Limnol. Oceanogr.* **27**, 184–188 (1982).
- Smol, J. P. The ratio of diatom frustules to chrysophytean statospores: a useful paleolimnological index. *Hydrobiologia* **123**, 199–208 (1985).
- Shemesh, A. *et al.* An oxygen isotope record of lacustrine opal from a European Maar indicates climatic stability during the last interglacial. *Geophys. Res. Lett.* **28**, 2305–2308 (2001).

Acknowledgements

The sediment core was retrieved thanks to the EU project Euromaar. This work was also supported by the TMR scheme operated by the EU, the UK NERC, the Israel Science Foundation and CNRS.

Correspondence and requests for materials should be addressed to P.R. (e-mail: prioual@geog.ucl.ac.uk).

An arbuscular mycorrhizal fungus accelerates decomposition and acquires nitrogen directly from organic material

Angela Hodge*, Colin D. Campbell† & Alastair H. Fitter*

* Department of Biology, University of York, PO Box 373, York, YO10 5YW, UK

† Macaulay Land Use Research Institute, Craigiebuckler, Aberdeen, AB15 8QH, UK

Arbuscular mycorrhizal fungi (order Glomales), which form mycorrhizal symbioses with two out of three of all plant species¹, are believed to be obligate biotrophs that are wholly dependent on the plant partner for their carbon supply². It is thought that they possess no degradative capability and that they are unable to decompose complex organic molecules, the form in which most soil nutrients occur. Earlier suggestions that they could exist saprotrophically were based on observation of hyphal proliferation on organic materials^{3,4}. In contrast, other mycorrhizal types have been shown to acquire nitrogen directly from organic sources^{5–7}. Here we show that the arbuscular mycorrhizal symbiosis can both enhance decomposition of and increase nitrogen capture from complex organic material (grass leaves) in soil. Hyphal growth of the fungal partner was increased in the presence of the organic material, independently of the host plant.

Nitrogen (N) is a critical limiting nutrient in many ecosystems⁸. Plants capture N largely in inorganic form, relying on microbes to release inorganic N (as NH_4^+) during decomposition of organic material. However, most N in soils is in organic form, often occurring in complex molecules. Some plants can take up simple, soluble organic N compounds^{9,10} and others can use organic N sources directly by association with specialist mycorrhizal fungi¹¹. To date, this mycorrhizal short cut in the N cycle has been shown only for ectomycorrhizas⁵ (about 5% of plant species, all woody) and ericoid⁷ mycorrhizas (about 1% of plant species). In contrast, fungi in the arbuscular mycorrhizal symbiosis promote plant growth principally by increasing uptake of immobile resources (notably inorganic phosphate), as they can acquire these beyond the depletion zone surrounding a root¹². They can also interfere with pathogens¹³, increase micronutrient uptake and alter drought resistance². It is thought that these fungi have little ability to increase plant uptake of more mobile ions (for example, NO_3^-) as these diffuse rapidly to roots in soil¹⁴, although they do transport the less mobile NH_4^+ (ref. 15). However, where mobile ions such as NO_3^- or NH_4^+ are being produced in decomposing patches of organic material, the ability of the fine hyphae of arbuscular mycorrhizal fungi to penetrate into the material and to compete with other microbes could lead to increased N acquisition by the plant. We tested the hypothesis that plant N uptake would be increased by giving the arbuscular mycorrhizal fungal hyphae access to a decomposing patch of organic matter from which roots were excluded.

We grew plants of *Plantago lanceolata* in microcosms (Fig. 1). Each microcosm comprised two rows of three compartments, four of which contained a single plant and two a patch of dual-labelled ($^{15}\text{N}/^{13}\text{C}$) *Lolium perenne* leaves containing 1.7 mg ^{15}N and 5.0 mg ^{13}C . The plant in the central compartment in each strip was inoculated with the mycorrhizal fungus *Glomus hoi*. The compartments containing plants were separated by a double layer of 20- μm mesh, which was permeable to hyphae but not roots—in one row (experimental) the compartment containing a patch of decomposing organic material (hereafter referred to as patch) was separated by a double 20- μm mesh on one side; in the other row (control)

separation was by a 0.45- μm mesh, through which hyphae cannot normally pass but which allows solute diffusion to occur.

Decomposition of the experimental patch was faster than the control: after 28 and 42 days (with no difference between these two sampling times) the mean ^{13}C content of the experimental patch samples was $0.21 \pm 0.08 \text{ mg g}^{-1}$ soil compared with $0.56 \pm 0.11 \text{ mg g}^{-1}$ for the controls ($F_{1,16} = 8.24$; $P = 0.011$), representing 4% and 11% retention, respectively. ^{15}N content showed a similar pattern ($F_{1,16} = 5.23$; $P = 0.036$) with 12% (experimental) and 21% (control) retention. However, the microbial community as described by the phospholipid fatty acid (PLFA) profile was unaffected by treatment. The increased rate of decomposition was therefore not due to qualitative changes in the microbial community caused by the fungi, but was probably a direct result of the presence of arbuscular mycorrhizal fungal hyphae.

The hyphae of *G. hoi* proliferated extensively within both the inoculated plant compartment and the experimental patch compartment (Fig. 2a), eventually reaching a similar density in each. In contrast, after 42 days there was less hyphal growth into the compartment that contained the uninoculated plant (Fig. 2a), and there was no difference between experimental and control rows in that compartment (Fig. 2b). The roots of the uninoculated plant were only weakly colonized in both experimental and control rows, 5% and 8% root length, respectively, compared with 47% and 37% for the inoculated plants. The preferential growth of hyphae into the organic patch rather than into the uninoculated plant compartment was unexpected: if the fungus acquires all its carbon (C) from plant hosts, it might be expected to benefit from maximizing the number of hosts it colonizes, which in this case would result in preferential growth into the uninoculated plant compartment. In contrast, growth into a patch of organic matter will potentially benefit its existing plant partner.

The benefit to the plant was significant in terms of N capture from the patch. Although experimental plants (in which the hyphae had access to the patch) were no different from controls in their dry weight, N content or N concentration, they captured three times as much patch N ($15 \pm 1.9\%$) as control plants ($5 \pm 1.2\%$; Fig. 3a). In

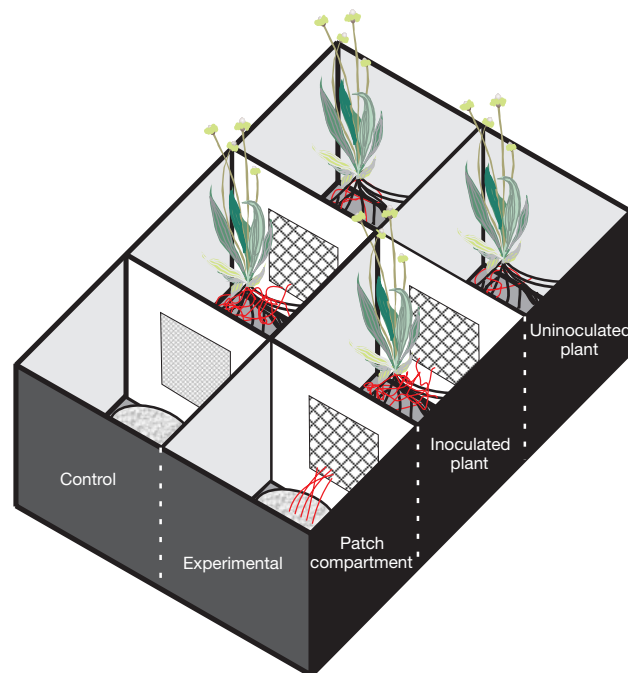


Figure 1 Microcosm unit design. The red lines in the compartments indicate arbuscular mycorrhizal hyphal mycelium; black lines indicate roots. Arbuscular mycorrhizal hyphae were denied access to the organic material on the control side of the microcosm unit.

contrast, uninoculated plants captured only 0.7% of patch N, irrespective of treatment. The ^{15}N in control plants must have been acquired by mass flow or diffusion; the additional N in experimental plants must have crossed the membrane that separated patch and plant compartments by means of arbuscular mycorrhizal fungal hyphae. The quantity of ^{15}N in the experimental plants was a simple function of hyphal length density in the patch (Fig. 3b), emphasising that arbuscular mycorrhizal fungal hyphae here had a role in N capture, and demonstrating that the enhanced ^{15}N capture by experimental plants was not simply due to an increased rate of diffusion across the membrane. There was no difference in shoot mass between experimental and control plants, which could have led to differences in convective water (and nitrate) flux; in addition, the double layer of mesh would minimize such fluxes. Arbuscular mycorrhizal fungal hyphae have been shown to transport N compounds in situations where diffusion and mass flow have been prevented by the use of hydrophobic barriers¹⁶.

The close relationship between N capture and hyphal density (Fig. 3b) is in marked contrast to the lack of a relationship between N uptake and root length density in other experiments^{17–19}. Even a low density of root length can capture all of a mobile resource in a patch; why therefore is this not true of hyphae? A possible explanation is that the hyphae were able to compete with other microbes for N released from the decomposing patch. As they are very fine (less

than 5 μm diameter), they can more easily penetrate throughout the material; more importantly, as they actively promoted decomposition—possibly by promoting the activity of hyphosphere bacteria—the hyphae must have been present at the sites of N release. Spatial proximity to the process of N release may make them effective scavengers for N.

The discovery that arbuscular mycorrhizal fungal hyphae may preferentially colonize organic patches rather than new host plants is surprising, even though an association between hyphal growth and organic matter is long established³. Current theory on the evolutionary strategy of the arbuscular mycorrhizal symbiosis²⁰ suggests that the best way for fungi to acquire C is to maximize the number of hosts that they colonize. In our data, the fungal growth pattern was more likely to enhance the performance of the existing host, even when no organic patch was made available, suggesting that promoting an existing host may be a more profitable strategy.

The increase in plant uptake of N from the patch demonstrates that arbuscular mycorrhizas may have a hitherto unappreciated role in plant N uptake in naturally heterogeneous soils where decomposing organic patches are crucial N sources²¹. In our experiment, the arbuscular mycorrhizal fungi both promoted decomposition and subsequently acquired the N that was released. They must also have used the products of decomposition, because they showed

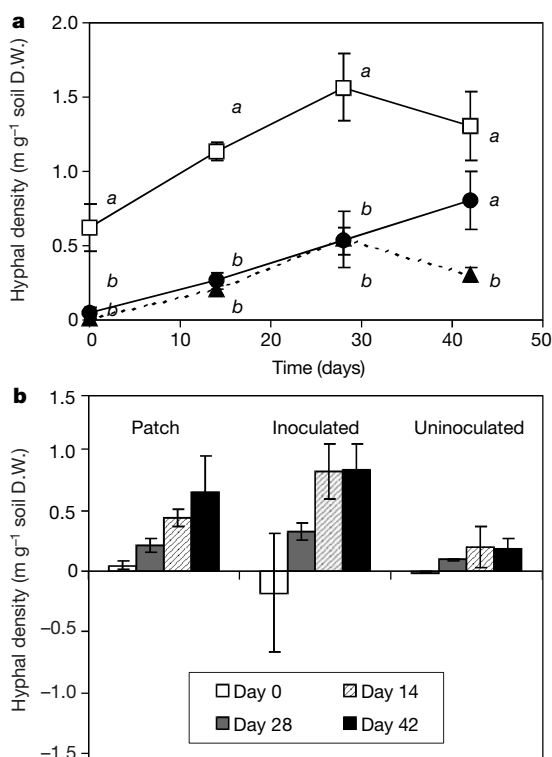


Figure 2 Arbuscular mycorrhizal hyphal densities. **a**, Hyphal densities in compartments of the experimental row of the microcosm. Inoculated plant, open squares; organic material, filled circles; uninoculated plant, filled triangles. Differences among compartments were significant ($F_{2,48} = 38.10$; $P < 0.001$). The letters beside data points represent significant differences at each collection date. Data were $\log_{10}$ transformed before statistical analysis. **b**, Difference between experimental and control values in each compartment. There was an increase in hyphal density when the arbuscular mycorrhizal fungus was allowed access to the organic material. Data are means with standard error bars. Analysis of variance showed that experimental and control rows differed in the patch ($F_{1,28} = 32.55$; $P < 0.001$) and inoculated plant ($F_{1,28} = 5.81$; $P = 0.023$) compartments, but not in the uninoculated plant compartment ($F_{1,28} = 0.82$; $P = 0.372$). D.W., dry weight.

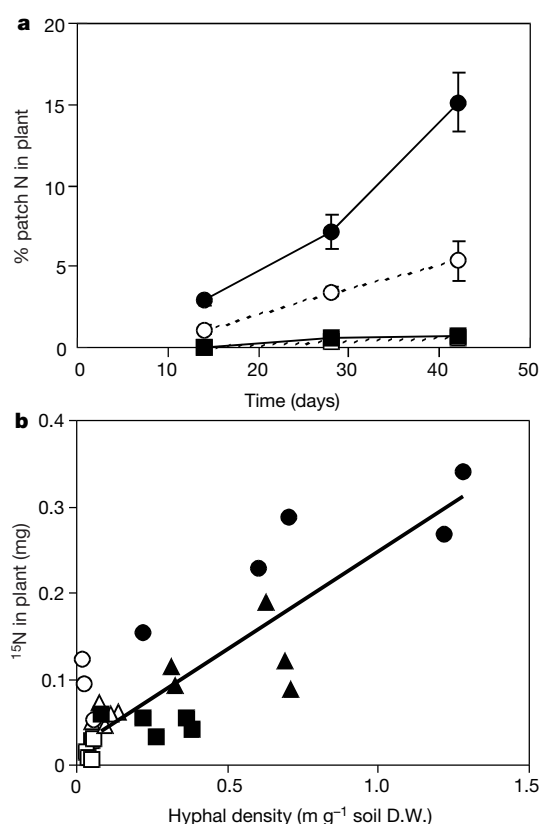


Figure 3 Nitrogen capture from the patch. **a**, Percentage N added in the organic patch detected in inoculated experimental plants (filled circles), inoculated control plants (open circles), uninoculated experimental plants (filled squares) and uninoculated control plants (open squares) over time. Data for uninoculated plants are superimposed. **b**, Relationship between the quantity of ^{15}N in the inoculated plant and the arbuscular mycorrhizal hyphal density in the organic patch compartment. There was no relationship for the control plant (open symbols) where arbuscular mycorrhizal hyphae were excluded from the patch. The experimental (filled symbols) plant data were fitted by a significant regression ($\text{mg } ^{15}\text{N} = 0.0208 + 0.226 \text{ m hyphae g}^{-1} \text{ soil}$; $P < 0.001$, $F_{1,13} = 24.23$, $r^2 = 65\%$). Data for each collection are distinguished by symbols (that is, day 14, squares; day 28, triangles; day 42, circles). D.W., dry weight.

markedly increased hyphal growth, both in the patch and in the inoculated plant compartment, despite the lack of any growth response by the plant partner in the symbiosis. These results show that the arbuscular mycorrhizal symbiosis can have saprotrophic capability, although the mechanism remains unknown and it is not yet established whether other arbuscular mycorrhizal fungi behave similarly. □

Methods

Microcosms

Twenty microcosm units were established under initially sterile conditions in containers measuring $39.5 \times 30 \times 16.3$ cm³, each divided into six compartments measuring $12 \times 13 \times 16$ cm³ and filled with a 1.8 kg mixture of autoclaved quartz sand and autoclaved (121 °C; 30 min) loam garden soil (1:1 v/v). The compartments were separated by double-thickness mesh barriers sealed to the sides with adhesive (see above). The two middle compartments of the microcosm unit each received 100 g inoculum of the arbuscular mycorrhizal fungus *G. hoi* (isolate number UY 110) as colonized roots of *P. lanceolata* in a sand and Terra-Green (a calcined attapulgite clay soil conditioner; Turf-Pro) growth medium. All compartments of the microcosm units received 50 ml filtered washing of the arbuscular mycorrhizal inoculum, but without arbuscular mycorrhizal propagules, to correct for differences in the starting microbial community^{22,23}. The experiment was set up in a randomized block design in a heated, lit glasshouse. The daily mean temperature over the duration of the experiment was 19.3 °C (standard error ± 0.08). Photosynthetically active radiation (PAR) flux was recorded weekly and averaged $464 \mu\text{mol m}^{-2} \text{s}^{-1}$ at plant level. All compartments of the microcosm units were watered daily. Four compartments of each of the microcosm units were planted with single *P. lanceolata* L. seedlings on 12 July 2000. The two remaining compartments received 0.55 g organic material 26 days later. The organic material added was milled *L. perenne* L. shoots that were dual-labelled¹⁹ with ¹³C and ¹⁵N such that 35 mg N (1.7 mg ¹⁵N) and 223 mg C (5.0 mg ¹³C) was added to each unplanted compartment. Grass leaves are typical of the range of organic materials added to soil in heterogeneous distributions. The material was milled to ensure that the organic material added in the patch compartment was uniform and homogenous. The organic material was added at a depth of 10 cm as a thin layer of material in a circle of diameter 6.5 cm and at a distance of 2.5 cm from the mesh to the patch perimeter. Units were collected 0, 14, 28 and 42 days after addition of the organic patch. At the final collection (42 days) roots from two of the plants in the control row had entered the patch by breaking through the adhesive that sealed the barrier between compartments to the box; the mesh was intact in all cases. These replicates are therefore excluded from the analysis. There were no significant differences in shoot, root or total dry weights between inoculated control and experimental plants and the uninoculated control and experimental plants. At day 42 the inoculated plant dry weight was 7.1 ± 0.41 g and the uninoculated plant dry weight was 7.7 ± 0.57 g.

Analyses

We used a subsample of root material for mycorrhizal assessment²⁴. Arbuscular mycorrhizal fungi hyphal length was measured using a modified membrane filter technique²⁵. Shoot, root and soil samples from the unplanted compartments were dried (70 °C; 48 h) and milled in a ball mill to a fine powder for analysis by mass spectrometry performed on a continuous-flow isotope ratio mass spectrometer (CF-IRMS). Data presented are for ¹³C and ¹⁵N in excess of the natural abundances of ¹⁵N and ¹³C. PLFA profiles were extracted using the modified Bligh and Dyer method²⁶ and analysed by gas chromatography²⁷. Individual PLFA, expressed as mol %, were analysed by multivariate analysis to determine changes in the overall microbial community structure. For statistical analysis, data were checked by a one-sample Kolmogorov–Smirnov test and transformed appropriately to ensure the data did not differ from a normal distribution before parametric analysis. Data were analysed using the general linear model (GLM) factorial design command in SPSS v. 7.0 for factorial analyses, or by one-way analysis of variance (ANOVA) or linear regression in Minitab v.12.1.

Received 23 May; accepted 9 August 2001.

1. Trappe, J. M. in *Ecophysiology of VA Mycorrhizal Plants* (ed. Safir, G.) 5–25 (CRC Press, Boca Raton, 1987).
2. Smith, S. E. & Read, D. J. *Mycorrhizal Symbiosis* (Academic, San Diego, 1996).
3. Mosse, B. Observations on the extramatrical mycelium of a vesicular-arbuscular endophyte. *Trans. Brit. Mycol. Soc.* **42**, 439–448 (1959).
4. St. John, T. V., Coleman, D. C. & Reid, C. P. P. Association of vesicular-arbuscular mycorrhizal hyphae with soil organic particles. *Ecology* **64**, 957–959 (1983).
5. Abuzinadah, R. & Read, D. J. The role of proteins in the nitrogen nutrition of ectomycorrhizal plants V. Nitrogen transfer in birch (*Betula pendula* L.) infected with different mycorrhizal fungi. *New Phytol.* **112**, 55–60 (1989).
6. Hodge, A., Alexander, I. J. & Gooday, G. W. Chitinolytic enzymes of pathogenic and ectomycorrhizal fungi. *Mycol. Res.* **99**, 935–941 (1995).
7. Leake, J. R. & Read, D. J. Proteinase activity in mycorrhizal fungi II. The effects of mineral and organic nitrogen sources on induction of extracellular proteinase in *Hymenoscyphus ericae* (Read) Korf and Kernan. *New Phytol.* **116**, 123–128 (1990).
8. Vitousek, P. M. & Howarth, R. W. Nitrogen limitation on land and in the sea: how can it occur? *Biogeochemistry* **13**, 87–115 (1991).
9. Chapin, F. S., Moilanen, L. & Kielland, K. Preferential use of organic nitrogen for growth by a non-mycorrhizal arctic sedge. *Nature* **361**, 150–153 (1993).

10. Jones, D. L. & Darrah, P. R. Re-sorption of organic components by roots of *Zea mays* L. and its consequences in the rhizosphere. I. Re-sorption of ¹⁴C labelled glucose, mannose and citric acid. *Plant Soil* **143**, 259–266 (1992).
11. Read, D. J. Mycorrhizas in ecosystems. *Experientia* **47**, 376–391 (1991).
12. Sanders, F. E. & Tinker, P. B. Mechanism of absorption of phosphate from soil by *Endogone mycorrhizas*. *Nature* **233**, 278–279 (1971).
13. Newsham, K. K., Fitter, A. H. & Watkinson, A. R. Multi-functionality and biodiversity in arbuscular mycorrhizas. *Trends Ecol. Evol.* **10**, 407–411 (1995).
14. Tinker, P. B. H. & Nye, P. H. *Solute Transport in the Rhizosphere* (Oxford Univ. Press, Oxford, 2000).
15. Tobar, R. M., Azcón, R. & Barea, J. M. The improvement of plant N acquisition from an ammonium-treated, drought-stressed soil by the fungal symbiont in arbuscular mycorrhizae. *Mycorrhiza* **4**, 105–108 (1994).
16. Mäder, P. *et al.* Transport of ¹⁵N from a soil compartment separated by a polytetrafluoroethylene membrane to plant roots via the hyphae of arbuscular mycorrhizal fungi. *New Phytol.* **146**, 155–161 (2000).
17. Van Vuuren, M. M. L., Robinson, D. & Griffiths, B. S. Nutrient inflow and root proliferation during the exploitation of a temporally and spatially discrete source of nitrogen in soil. *Plant Soil* **178**, 185–192 (1996).
18. Fransen, B., de Kroon, H. & Berendse, F. Root morphological plasticity and nutrient acquisition of perennial grass species from habitats of different nutrient availability. *Oecologia* **115**, 351–358 (1998).
19. Hodge, A., Stewart, J., Robinson, D., Griffiths, B. S. & Fitter, A. H. Root proliferation, soil fauna and plant nitrogen capture from nutrient-rich patches in soil. *New Phytol.* **139**, 479–494 (1998).
20. Bever, J. D. Dynamics within a mutualism and the maintenance of diversity: inference from a model of interguild frequency dependence. *Ecol. Lett.* **2**, 52–61 (1999).
21. Stark, J. M. in *Exploitation of Environmental Heterogeneity by Plants* (eds Caldwell, M. M. & Pearcy, R. W.) 255–284 (Academic, San Diego, 1994).
22. Koide, R. T. & Li, M. Appropriate controls for vesicular arbuscular mycorrhizal research. *New Phytol.* **111**, 35–44 (1989).
23. van der Heijden, M. G. A. *et al.* Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* **396**, 69–72 (1998).
24. Kormanik, P. P. & McGraw, A.-C. in *Methods and Principles of Mycorrhiza Research* (ed. Schenck, N. C.) 37–46 (American Phytopathol. Soc., St Paul, Minnesota, USA, 1982).
25. Staddon, P. L., Fitter, A. H. & Graves, J. D. Effect of elevated atmospheric CO₂ on mycorrhizal colonization, external mycorrhizal hyphal production and phosphorus inflow in *Plantago lanceolata* and *Trifolium repens* in association with the arbuscular mycorrhizal fungus *Glomus mosseae*. *Global Change Biol.* **5**, 347–358 (1999).
26. Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **37**, 911–917 (1959).
27. Frostegård, A., Tunlid, A. & Bååth, E. Changes in microbial community structure during long-term incubation in two soils experimentally contaminated with metals. *Soil Biol. Biochem.* **28**, 55–63 (1996).

Acknowledgements

We thank C. Scrimgeour and W. Stein for conducting the mass spectrometry analysis; P. Roberts for drawing Fig. 1; and D. Robinson and P. Young for comments on the manuscript. This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC).

Correspondence and requests for materials should be addressed to A.H. (e-mail: ah29@york.ac.uk).

The rhythm of microbial adaptation

Philip Gerrish

Theoretical Biology and Biophysics, T-10, Mailstop K710, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

The evolutionary biologist “studies the steps by which the miraculous adaptations so characteristic of every aspect of the organic world have evolved”¹. But the general nature of such adaptive steps is still unclear. Evolution is often thought to be random and dependent on unpredictable events². In this light, one might expect the steps taken by adaptation to be completely random, both biologically and temporally. Here I present a mathematical derivation to show that, on the contrary, adaptive steps can have fairly strong rhythm. I find that the strength of the adaptive rhythm, that is its relative temporal regularity, is equal to a constant that is the same for all microbial populations. As a consequence, numbers of accumulated adaptations are predicted to have a universal variance/mean ratio. The theory derived here is potentially applicable to the study of molecular evolution. Populations of organisms adapt to their environment through the

production of beneficial mutations and the subsequent spread of these mutations to predominance in the population through natural selection, a process loosely known as fixation. The fixation of a given beneficial mutation often takes a long time. During this time, it is possible, and likely in large microbial populations, that one or more alternative beneficial mutations will appear in the still-prevalent parental lineage in which the given beneficial mutation appeared. The result is the simultaneous presence in the population of several new lineages that each carry a selective advantage over their common progenitor. In microbial populations, genetic linkage (that is, scarcity of recombination) between beneficial mutations will cause a large fraction of these lineages to remain in direct competition throughout the ensuing contest for fixation. This period of competition introduces a degree of predictability in the time that elapses before a winning, or successful, lineage is fixed. The result is relative uniformity in the time intervals between fixations.

The phenomenon of competition among lineages created by beneficial mutations has been called the "Hill–Robertson effect" for sexual populations³ and "clonal interference" for asexual populations^{4,5}. It effects were deduced by Fisher⁶ and Muller⁷, and it was later modelled quantitatively^{5,8–11}. Historically, this phenomenon has most often been considered in the context of discussions^{6,7,12} and models^{13–15} of the evolutionary advantage of sex: it can be seen as a source of inefficiency in natural selection which can be remedied by increased recombination. Data from experiments with *Escherichia coli*¹⁶ and vesicular stomatitis virus^{4,17} confirm model predictions^{5,9} that, as mutation rate or population size increases, the number of competing lineages should increase and therefore (1) the fitness advantage of the winning lineage should increase and (2) the rate at which winning lineages are fixed should increase in a decelerating way. The temporal regularity of adaptation, however, was not addressed in either the experiments or previous models. Here I focus only on the temporal aspect of adaptation. Its relative simplicity reveals its generality.

Many beneficial mutations are lost owing to random sampling, or genetic drift¹⁸. During their brief existence, these mutations almost never achieve high enough frequency to affect the population fitness and are thus inconsequential from an evolutionary standpoint. Furthermore, a beneficial mutation may occur on a genome

containing one or more deleterious mutations. If the net fitness of the resulting mutant lineage is lower than the population mean fitness, then this beneficial mutation will almost certainly be lost.

The beneficial mutations of interest here are only those that survive genetic drift and are not linked to strongly deleterious mutations. Such mutations are here called contending mutations, because they are viable contenders for fixation. Eventually, a contending mutation is produced whose fitness advantage is large enough that it will not meet any superior competitors (that is, better contending mutations) on its way to fixation. This is called a successful mutation because it achieves fixation and thus evolutionary success. In what follows, the number of adaptations accumulated by a population is taken to be equal to the number of successful mutations that have been fixed.

Numbers of contending mutations produced between fixation events, N , have the moment-generating function derived in the Methods and given by equation (4). Standard usage of moment-generating functions¹⁹ directly gives the asymptotic coefficient of variation for N :

$$\frac{\sqrt{V(N)}}{E(N)} \xrightarrow{j} \sqrt{2e^{-\gamma} - 1} \quad (1)$$

where V is the true variance and E is the mean; $\gamma = 0.577$, Euler's constant. Increasing j may be interpreted as increasing population size or mutation rate, as these are correlated (see Methods). This value, roughly equal to one-third, defines the relative temporal regularity, or rhythmic strength, of adaptation and is independent of biological parameters.

What I have here called a 'rhythmic' process, indicating temporal regularity, is more rigorously defined in the probability literature and is there called a renewal process. In this language, adaptation may be thought of as a renewal process on the discrete time line formed by successive occurrences of contending mutations. This observation is useful for addressing the following question. Suppose

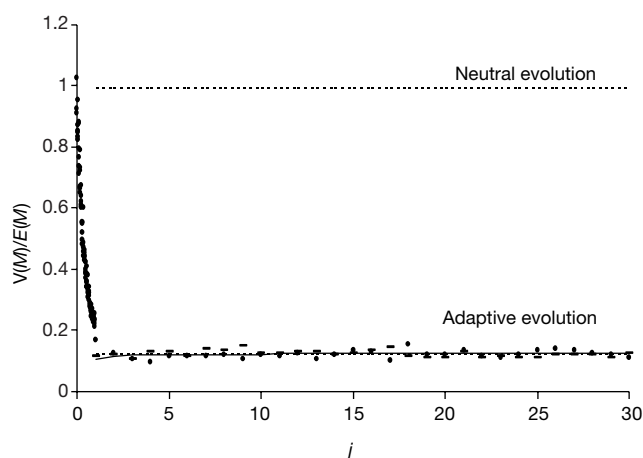


Figure 1 Variance/mean ratio for numbers of accumulated adaptations plotted against numbers of subsequent contending mutations, j . The solid line shows the exact solution; the lower dashed line (hardly distinguishable from solid line) shows the asymptote given by equation (2); the upper dashed line shows expectation for numbers of purely neutral changes. Dashes show the results of simulations of the simple process modelled by equation (3) with constant j . Dots show the results of simulations in which j is the average number of subsequent contending mutations and the appearance of contending mutations is a Poisson process.

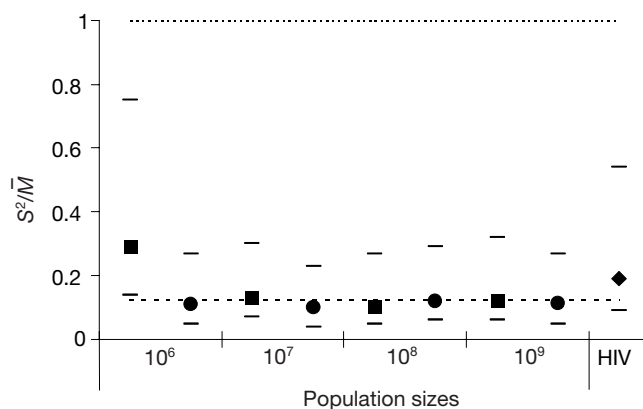


Figure 2 Comparing theory with simulations and HIV data (see Methods). Lower dashed line shows theoretical prediction for $V(M)/E(M)$ given by equation (2). Data points show values of $S^2/\bar{M}$, where $\bar{M}$ and S^2 are the sample mean and variance, respectively, for observed numbers of accumulated adaptations. The true variance/mean ratio (that is, $V(M)/E(M)$) lies between the error bars with 95% confidence. These calculations made the common assumptions that: (1) $nS^2/V(M)$ is $\chi^2(n-1)$ distributed, where n is sample size, S^2 is observed variance and $V(M)$ is true variance¹⁹, and (2) $\bar{M}$ is normally distributed with mean $E(M)$ and variance S^2/n . Population size categories show simulation results at different population sizes. Squares indicate that population size was constant; circles indicate that populations fluctuated by four orders of magnitude about the indicated population size; introducing a low level of recombination did not change results enough to merit plotting them separately. Beneficial mutations occurred at a rate of 5×10^{-9} per replication and had a mean selective advantage of 0.03. One in ten beneficial mutations was linked to a deleterious mutation with mean selective disadvantage of 0.03. Twenty simulations were run for each population size. The HIV category shows variance/mean ratio derived from HIV sequence data (see Methods).

several populations evolve independently and, at time t , I count the numbers of accumulated adaptations, M , in each population. How does the temporal regularity of adaptation affect these numbers? The answer comes from renewal theory: a central limit theorem²⁰ predicts that the random variable M will become normally distributed as time progresses, with

$$\frac{V(M)}{E(M)} \xrightarrow{t} \frac{V(N)}{[E(N)]^2} \xrightarrow{j} 2e^{-j} - 1 \approx 0.123 \quad (2)$$

This asymptotic variance/mean ratio for accumulated adaptations is again independent of biological parameters. The implication is that it applies equally well to any large population of adapting entities with a significant degree of genetic linkage (or analogue thereof). The rapid convergence of this equation with increasing j is illustrated in Fig. 1.

The comparison between neutral and adaptive evolution, in Fig. 1, raises the question of how the transition between these two might occur as a function of selection. When beneficial mutations confer extremely small selective advantages (less than the reciprocal of the population size²¹), their fate is determined mostly by unbiased random sampling. Their minuscule selective advantage has no observable effect on their trajectory and they are thus effectively neutral mutations. Fixations of such mutations are stochastic, forming a Poisson process²², and the numbers of such fixations over a given period of time therefore have a variance/mean ratio equal to 1. As their selective advantage increases, it begins to affect their fate, and beneficial mutations begin to compete. In other words, contending mutations begin to appear, meaning that j becomes positive. As shown in Fig. 1, the transition from neutral to adaptive evolution is very rapid with increasing j , mostly occurring in the region $0 < j < 1$. As selective advantage increases, therefore, a rapid transition occurs, through j , from the purely neutral case to the adaptive case, and the variance/mean ratio jumps (continuously) from 1 to 0.123 at very small selective advantages.

In Fig. 2, equation (2) is compared with stochastic simulations of evolving bacterial populations under various conditions (see Methods). The average number of accumulated adaptations in the simulations was only 4.8, indicating that convergence of equation (2) occurs in relatively short evolutionary time. The HIV category in Fig. 2 shows how some biological data correspond to theory: non-synonymous changes (that is, mutations that alter the amino-acid composition of the gene product) in the third variable region of the envelope gene from HIV-1 were analysed (see Methods). I chose this region because it is known to be exposed to the host immune system, meaning that most non-synonymous changes should affect fitness.

In addition to the good agreement with simulations and viral data, two other issues argue strongly for the robustness of equation (2). The first is the fact that its progenitor, equation (3) in the Methods, is independent of the fitness distribution for lineages created by contending mutations. Many potential confounding factors can be thought of as effectively altering this fitness distribution, factors such as: (1) competition among double and triple mutants in very large populations²³, (2) linkage between beneficial and deleterious mutations²⁴, or (3) a change in the fitness landscape, and thus the fitness distribution²⁵, with each fixation. But because of their independence of the fitness distribution, the results derived here remain unchanged by such factors. The second issue is that equation (2) is nearly independent of how numbers of subsequent contending mutations are distributed (see Methods). Fluctuations in these numbers can arise as a result of (1) fluctuations in population size, or (2) fluctuations in the availability of potential beneficial mutations owing to varying degrees of adaptiveness. But because of the second issue mentioned above, the results derived here are affected very little by such fluctuations (see Fig. 2).

One area of application of the results derived here is molecular evolution and the interpretation of sequence data. To give an example, estimates of divergence times and corresponding confidence intervals could be based on the renewal process of adaptation, derived here, where appropriately scaled times between non-synonymous changes have a moment-generating function given by equation (4). These estimates would complement estimates based on the traditional Poisson assumption that is reflective of neutral evolution. Also, the theory derived here could provide a theoretical framework for explicit modelling of molecular evolution, where 'hitchhiking' of neutral and deleterious mutations is taken into account. Empirical evidence and simulation evidence, together with strong theoretical arguments for robustness, support the general applicability of the results derived here. □

Methods

Derivation of moment-generating function

Let the fitnesses of lineages created by contending mutations be independent, identically distributed random variables drawn from an unknown distribution. (This assumption for fitnesses need only hold for the duration of one fixation event, a fact that is implicit in this derivation.) Suppose that during the time required for its fixation, a contending mutation meets exactly j competitors (see below for relaxation of the assumption of constant j). Then the successful mutation in a given parental lineage is precisely defined as the first contending mutation to be superior to all previous and j subsequent such mutations. The probability that the i th contending mutation is the successful one is equal to the probability that (1) it is the best of $i + j$ contending mutations, and (2) no candidate for success was previously fixed in the same parental lineage. Condition (2) requires clarification: if the i th contending mutation is the best of the first $i + j$, this necessarily implies that contending mutations $i - j, i - j + 1, \dots, i - 1$ cannot have been successful, because of how success has been defined. But it does not necessarily exclude the possibility that the $i - j - 1$ th or a previous contending mutation was successful, and the probabilities of success for these earlier contending mutations must therefore be discounted. The probability that the i th contending mutation is the successful one is thus given by the product

$$p(i|j) = \frac{1}{i+j} \left(1 - \sum_{k=1}^{i+j-1} p(k|j) \right) \quad (3)$$

These probabilities are completely independent of the distribution of fitnesses of lineages created by contending mutations. Equation (3) has an alternative form: $(i + j)p(i|j) = (i + j - 1)p(i - 1|j) - p(i - j - 1|j)$. Multiplying this equation by z^i , where z is a dummy variable, and summing from $i = 0$ to ∞ puts the equation in terms of $P(z|j) = \sum_{i=0}^{\infty} p(i|j)z^i$, the probability-generating function for the $p(i|j)$, and derivatives thereof. The zeroth contending mutation does not exist and so cannot be fixed. The first contending mutation will be the successful mutation (that is, it will be fixed) if it is the best of $j + 1$ contending mutations. These two observations give the first two probabilities, $p(0|j) = 0$ and $p(1|j) = 1/(j + 1)$. Solving the resulting differential equation for $P(z|j)$ gives $P(z|j) = z^{-j}[1 + (z - 1)\exp\{\sum_{k=1}^j (z^k/k)\}]$, the probability-generating function for numbers of contending mutations up to and including the successful one. The probability-generating function for total numbers of contending mutations produced between fixation events, $N = i + j$, is obtained by multiplying this probability-generating function by z^j , giving $\tilde{P}(z|j) = 1 + (z - 1)\exp\{\sum_{k=1}^j (z^k/k)\}$. The corresponding moment-generating function is then simply $Q(s|j) = \tilde{P}(e^s|j)$. Thus, the random variable N is shown to have moment-generating function

$$Q(s|j) = 1 + (e^s - 1)\exp\left\{\sum_{k=1}^j \frac{e^k}{k}\right\} \quad (4)$$

It is readily verified that equation (3) generates a probability density by noting that $Q(0|j) = 1$, for all j .

Relaxing the assumption of constant j

For simplicity of presentation, the number of subsequent contending mutations, j , has been assumed to be constant. This assumption may seem an over-simplification, yet its relaxation leaves equation (2) unchanged. If constant j is replaced with random variable J , then it can be shown that equation (2) holds for all distributions for which $[2V(J) + E(J)]/[E(J)]^2 \xrightarrow{B(J)} 0$. Furthermore, extensive simulations of the simple process modelled by equation (3) with different distributions for J have shown equation (2) to be very robust and have always converged rapidly. Under the reasonable assumption that contending mutations occur as a Poisson process, simulations show very fast convergence (Fig. 1). These simulations are distinct from and complement the more complex simulations described below.

Simulations of bacterial populations

These simulations track each lineage created by beneficial mutation. A lineage starts with a single mutant individual and grows stochastically until it acquires 100 members, after

which its growth is deterministic. (Of course, many lineages quickly die out owing to their stochastic trajectory, or genetic drift, and thus never acquire 100 members.) Let f_i denote the frequency of lineage i , and let μ denote beneficial mutation rate. Each generation, a Poisson-distributed number of new beneficial mutations arises in each lineage with mean μf_i forming new lineages. The new beneficial mutations confer selective advantages that are drawn from an exponential distribution. There is a ten per cent chance that a new beneficial mutation is linked to a deleterious mutation, whose selective disadvantage is drawn from an exponential distribution. The program tracks the cumulative number of beneficial mutations acquired by a lineage. After a prescribed number of generations, the lineage that has the highest frequency is determined and its cumulative number of beneficial mutations is recorded. This number is taken to be the number of adaptations accumulated by the population in the prescribed number of generations. The simulation code is freely available upon request.

Received 27 February; accepted 23 July 2001.

- Mayr, E. *Toward a New Philosophy of Biology* 26 (Harvard Univ. Press, Cambridge, Massachusetts, 1988).
- Gould, S. J. *Wonderful Life: the Burgess Shale and the Nature of History* (W. W. Norton, New York, 1989).
- Hill, W. G. & Robertson, A. The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).
- Miralles, R., Gerrish, P. J., Moya, A. & Elena, S. F. Clonal interference and the evolution of RNA viruses. *Science* **285**, 1745–1747 (1999).
- Gerrish, P. J. & Lenski, R. E. The fate of competing beneficial mutations in an asexual population. *Genetica* **102–103**, 127–144 (1998).
- Fisher, R. A. *The Genetical Theory of Natural Selection* 102–104, 121–123 (Oxford Univ. Press, Oxford, 1930).
- Muller, H. J. Some genetic aspects of sex. *Am. Nat.* **8**, 118–138 (1932).
- McVean, G. A. T. & Charlesworth, B. The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics* **155**, 929–944 (2000).
- Haigh, J. The accumulation of deleterious genes in a population—Muller's ratchet. *Theor. Pop. Biol.* **14**, 251–267 (1978).
- Robertson, A. in *Mathematical Topics in Population Genetics* (ed. Kojima, K.) 246–288 (Springer, Berlin, 1970).
- Barton, N. H. The reduction in fixation probability caused by substitutions at linked loci. *Genet. Res.* **64**, 199–208 (1994).
- Morgan, T. H. *Heredity and Sex* (Columbia Univ. Press, New York, 1913).
- Felsenstein, J. The evolutionary advantage of recombination. *Genetics* **78**, 737–756 (1974).
- Barton, N. H. Linkage and the limits to natural selection. *Genetics* **140**, 821–841 (1995).
- Otto, S. P. & Barton, N. H. The evolution of recombination: Removing the limits to natural selection. *Genetics* **147**, 879–906 (1997).
- de Visser, J. A. G. M., Zeyl, C. W., Gerrish, P. J., Blanchard, J. L. & Lenski, R. E. Diminishing returns from mutation supply rate in asexual populations. *Science* **283**, 404–406 (1999).
- Miralles, R., Moya, A. & Elena, S. F. Diminishing returns of population size in the rate of RNA virus adaptation. *J. Virol.* **74**, 3566–3571 (2000).
- Haldane, J. B. S. The mathematical theory of natural and artificial selection. *Proc. Camb. Phil. Soc.* **23**, 838–844 (1927).
- Hogg, R. V. & Craig, A. T. *Introduction to Mathematical Statistics* 59–67, 276 (Prentice Hall, Upper Saddle River, New Jersey, 1995).
- Feller, W. *An Introduction to Probability Theory and Its Applications* Vol. 2, 346–359 (John Wiley & Sons, New York, 1968).
- Gillespie, J. H. *Population Genetics: A Concise Guide* 77–82 (John Hopkins Univ. Press, Baltimore, 1998).
- Ohta, T. & Kimura, M. On the constancy of the evolutionary rate of cistrons. *J. Mol. Evol.* **1**, 18–25 (1971).
- Wahl, L. M. & Krakauer, D. C. Models of experimental evolution: the role of genetic chance and selective necessity. *Genetics* **156**, 1437–1448 (2000).
- Orr, H. A. The rate of adaptation in asexuals. *Genetics* **155**, 961–968 (2000).
- Orr, H. A. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* **52**, 935–949 (1998).
- The WHO Network For HIV Isolation And Characterization. HIV type 1 variation in World Health Organization-sponsored vaccine evaluation sites: Genetic screening, sequence analysis, and preliminary biological characterization of selected viral strains. *AIDS Res. Hum. Retroviruses* **10**, 1327–1343 (1994).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Felsenstein, J. PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164–166 (1989).
- Korber, B. T. SNAP: Synonymous-Nonsynonymous Analysis Program. Available at <http://www.hiv.lanl.gov>. (Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, New Mexico, 1998).

Acknowledgements

I thank C. Macken, T. Johnson, P. Sniegowski, A. Orr, D. Chench, J. Gerrish, P. Baccam, R. Lenski and D. Krakauer for discussions. This work was supported in part by the Postdoctoral Fellowship program at Los Alamos National Laboratories and in part by the Laboratory Directed Research and Development program through financial support to C. Macken at Los Alamos National Laboratories.

Correspondence and requests for materials should be addressed to the author (e-mail: gerrish@lanl.gov).

An agrin minigene rescues dystrophic symptoms in a mouse model for congenital muscular dystrophy

Joachim Moll*, Patrizia Barzaghi*, Shuo Lin*, Gabriela Bezakova*, Hanns Lochmüller†, Eva Engvall‡, Ulrich Müller§ & Markus A. Ruegg*

* Department of Pharmacology/Neurobiology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

† Genzentrum Munich, Ludwig-Maximilians-University, Feodor-Lynen-Strasse 25, D-81377 Munich, Germany

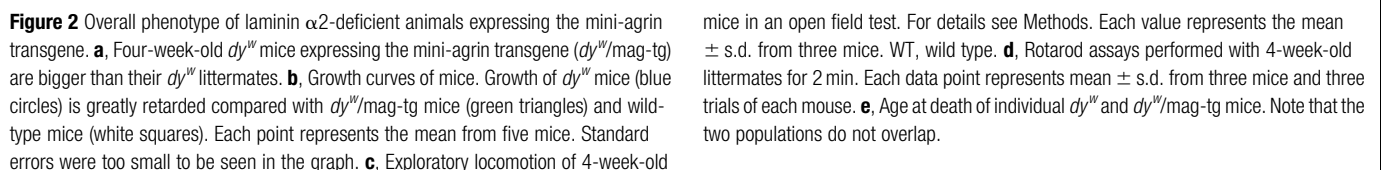
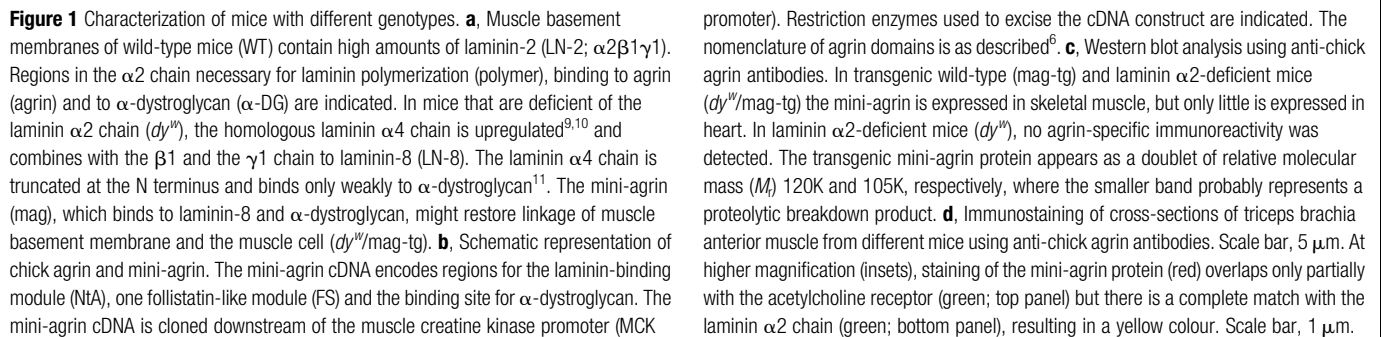
‡ The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, USA

§ Friedrich-Miescher Institute, Maulbeerstrasse 66, CH-4058 Basel, Switzerland

Congenital muscular dystrophy is a heterogeneous and severe, progressive muscle-wasting disease that frequently leads to death in early childhood^{1,2}. Most cases of congenital muscular dystrophy are caused by mutations in *LAMA2*, the gene encoding the $\alpha 2$ chain of the main laminin isoforms expressed by muscle fibres. Muscle fibre deterioration in this disease is thought to be caused by the failure to form the primary laminin scaffold, which is necessary for basement membrane structure³, and the missing interaction between muscle basement membrane and the dystrophin–glycoprotein complex (DGC)⁴ or the integrins⁵. With the aim to restore muscle function in a mouse model for this disease, we have designed a minigene of agrin, a protein known for its role in the formation of the neuromuscular junction⁶. Here we show that this mini-agrin—which binds to basement membrane⁷ and to α -dystroglycan⁸, a member of the DGC—amends muscle pathology by a mechanism that includes agrin-mediated stabilization of α -dystroglycan and the laminin $\alpha 5$ chain. Our data provides *in vivo* evidence that a non-homologous protein in combination with rational protein design can be used to devise therapeutic tools that may restore muscle function in human muscular dystrophies.

Laminins are cross-shaped heterotrimers of α -, β - and γ -chains (see Fig. 1a). The five α -, three β - and three γ -chains give rise to at least twelve different protein isoforms (laminin-1 to laminin-12; reviewed in ref. 3), which differ in their tissue distribution⁹. Mutations in the laminin $\alpha 2$ chain, which is the most predominant α -chain in muscle and peripheral nerve, cause a most severe form of congenital muscular dystrophy, commonly referred to as merosin-deficient congenital muscular dystrophy (MCMD). Hallmarks of MCMD are a complete or partial loss of structured muscle basement membrane, ongoing muscle necrosis (which results in elevated levels of creatine kinase in the blood) and demyelination of peripheral nerves^{1,2}. Deficiency of laminin $\alpha 2$ is accompanied by upregulation of the laminin $\alpha 4$ chain, giving rise to laminin-8 (Fig. 1a; see also refs 9, 10). This α -chain, however, lacks domains that are important for laminin polymerization³, and binds only weakly to α -dystroglycan¹¹, a principal receptor for the $\alpha 2$ chain (Fig. 1a). We therefore proposed that MCMD pathology might be rescued by introducing an extracellular component that links laminin-8 to the muscle cell surface. Specifically, we postulated that expression of a miniaturized agrin molecule, which retained high-affinity binding sites for the laminins^{7,12} and α -dystroglycan⁸, might compensate the loss of the laminin $\alpha 2$ chain (Fig. 1a).

The heparan sulphate proteoglycan agrin is best known for its activity to induce postsynaptic specializations at the neuromuscular junction (reviewed in ref. 6). Alternative messenger RNA splicing generates protein isoforms that differ in their activity to induce postsynaptic differentiation. The minigene used in our studies is derived from an agrin isoform (termed muscle agrin) that is inactive



in inducing postsynaptic specializations both *in vitro* and *in vivo*^{6,13}. This agrin isoform is detected in muscle basement membranes during embryogenesis. In the adult, its principal sites of expression are kidney and lung⁶. We used a chick mini-agrin construct that comprised the amino-terminal agrin (NtA; ref. 12) domain, which binds to laminins^{7,12}, the first follistatin-like domain and the α -dystroglycan-binding domain (Fig. 1b). Expression of this complementary DNA in transgenic mice was driven by a 1.3-kilobase (kb) fragment of the mouse muscle creatine kinase (MCK) promoter (Fig. 1b; see also refs 14, 15). The mini-agrin transgenic (mag-tg) mice were then bred with animals carrying a null mutation in the *lama2* gene (*dy*^w; ref. 16). Homozygous *dy*^w mice show the same pathology as the spontaneous *lama2*-mutant mice (*dy*), which are the classical animal model for MCMD^{17,18}. Levels of the mini-

agrin protein were high in skeletal muscle but only small amounts were detected in the heart, independent of whether the transgenic mice also expressed the laminin α 2 chain or not (Fig. 1c). Anti-chick agrin antibodies stained the entire circumference of muscle fibres in mag-tg and *dy*^w/mag-tg (Fig. 1d, top panel) but not in *dy*^w or wild-type mice (Fig. 1d, bottom panel). Chick agrin-like immunoreactivity (red colour in Fig. 1d insets) only partially overlapped with acetylcholine receptors (top inset; green) but co-distributed with the laminin α 2 chain (bottom inset; green), indicating that the mini-agrin protein is an integral part of the muscle basement membrane. The staining pattern of the transgenic agrin protein was similar in all muscles examined, which included extensor digitorum longus, soleus, diaphragm and intercostal muscle (data not shown).

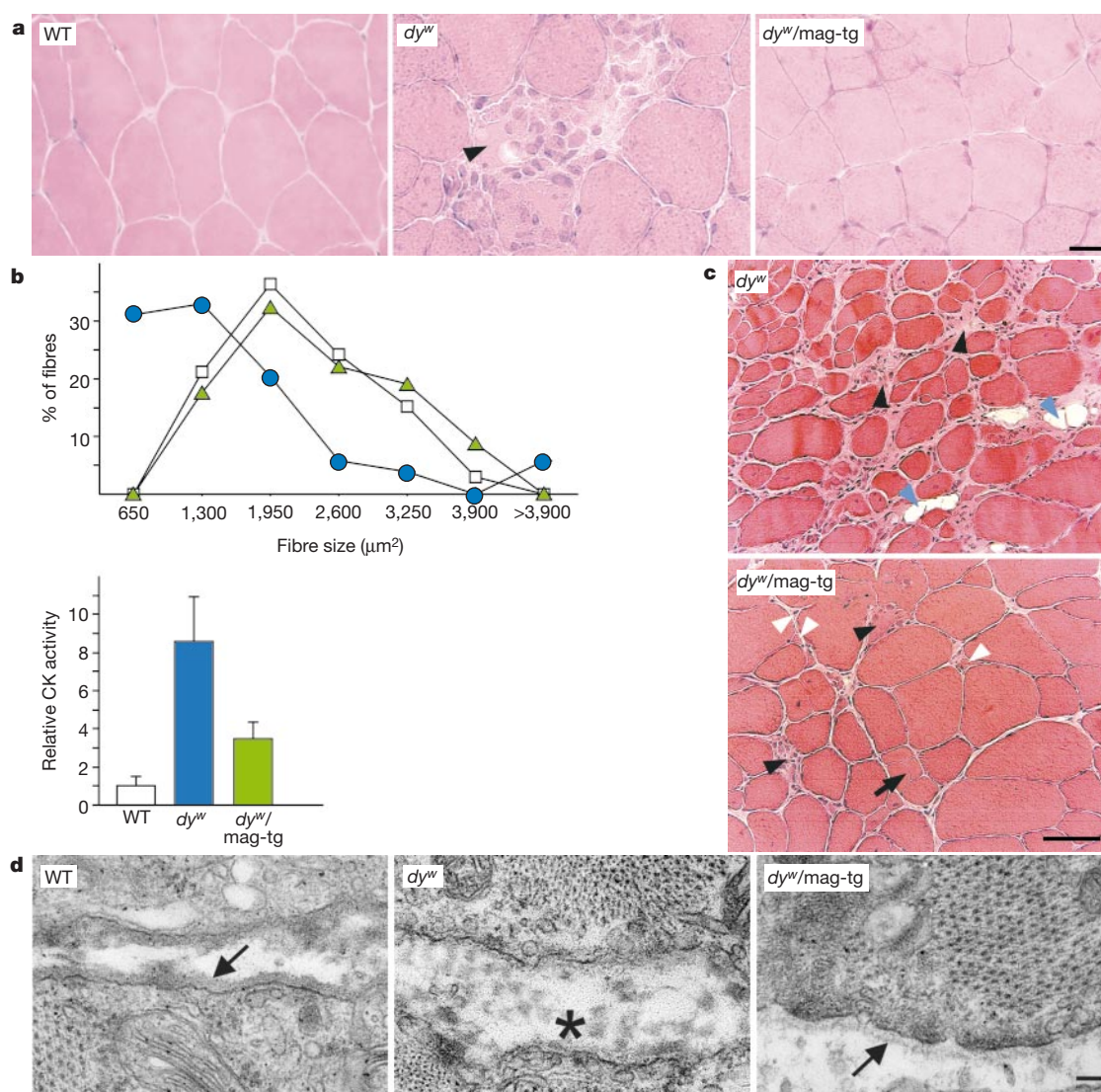


Figure 3 Analysis of muscle in the different mice. **a**, Haematoxylin and eosin staining of tibialis anterior muscles of 4-week-old mice. Arrowhead indicates degenerating muscle fibres in *dy*^w mice. Scale bar, 25 μm . WT, wild type. **b**, Quantification of muscle fibre size in 4-week-old mice (top). Muscle fibres from four different mice of each genotype (five sections each) were grouped into different classes and the relative percentage of fibres was determined for each class. The values given indicate the upper limit of each group. Laminin α 2-deficient *dy*^w mice (blue circles) have many very small and very large fibres. The distribution of muscle fibre size is more homogenous in wild-type (white squares) and *dy*^w/mag-tg (green triangles) mice. Creatine kinase (CK) activity in blood serum of 4-week-old animals (bottom). Values given reflect the mean $\pm$ s.d. of 12 (WT), 6 (*dy*^w) or 13

(*dy*^w/mag-tg) mice. For details see Methods. **c**, Haematoxylin and eosin staining of quadriceps muscle from 16-week-old mice. The severe myopathy in *dy*^w mice is indicated by the presence of fibrous (black arrowheads) and fatty (blue arrowheads) tissue. A mild, residual myopathy is detected in *dy*^w/mag-tg mice indicated by some degenerating muscle fibres (black arrowheads) and occasional fibre splitting (black arrow). Small polygonal muscle fibres (white arrowheads) are signs of an additional neurogenic lesion. Scale bar, 25 μm . **d**, Electron micrographs of triceps brachii from 4-week-old mice reveal that the electron-dense structure of the muscle basement membrane (arrows) is restored in *dy*^w/mag-tg mice. In *dy*^w mice, the basement membrane appears patchy (asterisk). Scale bar, 3 nm.

As reported earlier¹⁶, *dy^w* mice appear much more passive, smaller and thinner than wild-type mice. As shown in Fig. 2, the overall health of *dy^w/mag-tg* mice was greatly improved compared with their *dy^w* littermates. The weight gain and growth curves, which are greatly delayed in *dy^w* mice, were similar for *dy^w/mag-tg* and wild-type mice (Fig. 2a, b). The beneficial influence of the mini-agrin also became apparent in the locomotory activity of the mice. Locomotion was examined in an open field test by placing mice into a new cage and measuring the time that they spent walking around in the first 5 min. Wild-type mice spent more than 90% of the time in motion, whereas *dy^w* mice spent less than 20% of the time moving (Fig. 2c). This inertia was not seen in *dy^w/mag-tg* mice (Fig. 2c). Similar results were also apparent in a rotarod test. In this test, *dy^w* mice fell from the rotarod immediately, whereas most *dy^w/mag-tg* mice managed to stay on the rotarod (Fig. 2d). The beginning of partial hindleg paralysis in *dy^w/mag-tg* mice is probably the reason for their weaker performance compared with wild-type mice. This partial loss of motor control in *dy^w/mag-tg* mice is due to progressive demyelination of peripheral nerves¹⁶, a phenotype that is not rescued because the mini-agrin transgene is not expressed in peripheral nerves (data not shown). Expression of the agrin transgene also had a strong impact on the longevity of the mice (Fig. 2e). On average, *dy^w* mice died at the age of 8 weeks compared with around 40 weeks for *dy^w/mag-tg* mice.

Next we examined different muscles from 4-week-old mice by haematoxylin and eosin staining. All the hallmarks of MCMD were observed in *dy^w* mice, such as extensive degeneration and some regeneration of muscle fibres and an accompanied expansion of connective tissue (Fig. 3a; *dy^w*). In contrast, no degeneration was apparent in *dy^w* mice that expressed mini-agrin (Fig. 3a; *dy^w/mag-tg*). A typical sign of the ongoing degeneration and regeneration of muscle fibres is a variation in fibre size¹. A high percentage of very small (up to 1,300 μm^2) and some very large muscle fibres (greater than 3,900 μm^2) were found in *dy^w* mice. In contrast, *dy^w/mag-tg* mice displayed a more uniform distribution of fibre size that was indistinguishable from that in wild-type mice (Fig. 3b, top panel). The ongoing muscle necrosis was also manifested by the more than eight times higher creatine kinase levels in the blood of 4-week-old *dy^w* mice (Fig. 3b, bottom). In *dy^w* mice expressing mini-agrin (*dy^w/mag-tg*), creatine kinase activity was more than two times lower, but remained approximately three times higher than in wild-type mice (Fig. 3b, bottom panel). Creatine kinase activity is also approximately three times higher in the blood of *dy^w* mice that express a transgene of the human laminin $\alpha 2$ chain¹⁶, which is expected to fully rescue muscle pathology. Thus, the still elevated creatine kinase levels in *dy^w/mag-tg* mice may originate from heart muscle, from changes caused by the demyelination of peripheral nerves, or from residual myopathy that is not fully rescued by the mini-agrin transgene. To investigate this, we also examined haematoxylin- and eosin-stained muscle cross-sections of 16-week-old mice (Fig.

3c). The few *dy^w* mice that survived for 16 weeks showed signs of a chronic and severe degenerative myopathy. In these animals, fibrous (black arrowheads) and fatty (blue arrowheads) tissue had replaced a high percentage of the muscle, and most of the remaining fibres were round and small. In contrast, most of the muscle fibres in *dy^w/mag-tg* mice were of polygonal shape and normal size. Several small, angular muscle fibres, scattered throughout the muscle (white arrowheads), indicated a neurogenic lesion. In addition, myopathic symptoms became visible, which included a few degenerating fibres (black arrowheads) and some fibre splitting (black arrow). As shown in Fig. 3d, the mini-agrin also restored the structure of muscle basement membrane (arrow; *dy^w/mag-tg*), which is

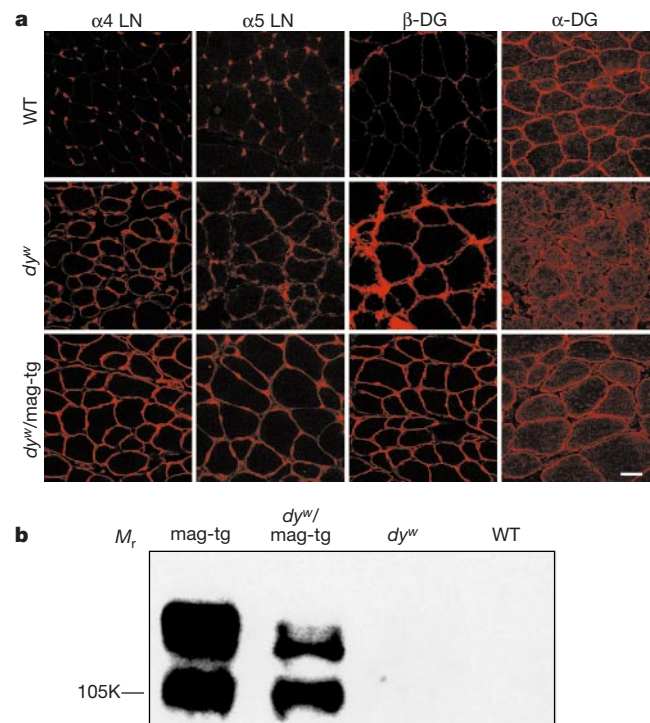


Figure 4 Involvement of laminin α chains and dystroglycan. **a**, Cross-sections of triceps brachii muscle were stained with antibodies directed against the proteins indicated at the top of each column. Cross-sections of all genotypes (indicated to the left of each row) were stained in parallel using conditions that guaranteed linearity of the staining intensity. WT, wild type; LN, laminin; DG, dystroglycan. Scale bar, 40 μm . **b**, The transgenic agrin protein co-purifies with α -dystroglycan. α -dystroglycan and its associated proteins were enriched from muscle homogenates using WGL Sepharose beads. After SDS-PAGE and western blotting nitrocellulose membranes were stained with anti-chick agrin antibodies. For details see Methods.

Table 1 Amount of protein detected in the different genotypes

Protein	<i>dy^w</i>		Western blot	<i>dy^w/mag-tg</i>		Western blot
	Confocal microscopy	Western blot		Confocal microscopy	Western blot	
Perlecan	106 $\pm$ 25	94 $\pm$ 23	110 $\pm$ 30	115 $\pm$ 17	108 $\pm$ 13	180 $\pm$ 30
Laminin $\gamma 1$	107 $\pm$ 27	104 $\pm$ 13	ND	117 $\pm$ 26	110 $\pm$ 14	ND
Nidogen	ND	ND	70 $\pm$ 10	ND	ND	90
Laminin $\alpha 4$	970 $\pm$ 100*	1,480 $\pm$ 200*	320 $\pm$ 110	1,090 $\pm$ 170*	1,810 $\pm$ 260*†	410 $\pm$ 30
Laminin $\alpha 5$	147 $\pm$ 29*	170 $\pm$ 40*	126 $\pm$ 72	440 $\pm$ 68*†	636 $\pm$ 79*†	371 $\pm$ 130
$\alpha 7$ integrin	157 $\pm$ 38*	100 $\pm$ 20	80 $\pm$ 20	181 $\pm$ 61	104 $\pm$ 28	90 $\pm$ 4
α -dystroglycan	33 $\pm$ 7*	22 $\pm$ 5*	NA	86 $\pm$ 14†	86 $\pm$ 15†	NA
β -dystroglycan	300 $\pm$ 77*	240 $\pm$ 30*	250 $\pm$ 90	350 $\pm$ 76*	250 $\pm$ 34*	200 $\pm$ 90

The amount of proteins detected in triceps brachii muscle of different genotypes using quantitative immunostaining (confocal microscopy) and immunoblotting (western blot). Values are expressed as percentages of the levels found in wild-type mice (normalized to 100%). In confocal microscopy measurements the two series of experiments are represented; the values are the mean $\pm$ s.e. ($N = 20$). For details see Methods. For immunoblotting experiments (western blot), data from two to five different mice were pooled and the values given represent the mean $\pm$ s.d. Because the values of the confocal microscopy measurements are not normally distributed, statistical analysis was performed with each series separately. Because of the low number of samples and non-normal distribution, data from Western blots were not subjected to statistical analysis. NA, not applicable; ND, not determined.

* Values significantly different from wild-type mice (Wilcoxon test, $P < 0.001$).

† Values significantly different from *dy^w* mice (Wilcoxon test, $P < 0.001$).

disrupted in *dy^w* mice¹⁹ (asterisk; *dy^w*), to the extent that it became indistinguishable from that of wild-type mice.

The proposed molecular mechanisms that underlie the pathology of laminin $\alpha 2$ -deficient MCMD include perturbation of interactions between the laminin $\alpha 2$ chain and its cell-surface receptors, dystroglycan and/or the integrins^{4,5}, and also the failure to form a laminin-based primary scaffold necessary for basement membrane integrity. The failure to form the laminin scaffold seems also to be the reason for the muscle dystrophy in the *dy²¹* mouse mutant. These mice express an amino-terminally-truncated version of the laminin $\alpha 2$ chain that is unable to form laminin polymers *in vitro* but still binds to α -dystroglycan and the integrins²⁰. The muscle dystrophy in *dy²¹* mutants²¹ is less severe than in mice that do not express the laminin $\alpha 2$ chain at all, suggesting that both mechanisms contribute to the severity of the disease in *dy^w* mutants. We measured the expression levels of different extracellular matrix components and their cellular receptors in mice of different genotype by western blot analysis and immunostaining. No changes were detected in the amounts of the extracellular matrix protein perlecan, nidogen and the laminin $\gamma 1$ chain (Table 1). Similarly, the levels of $\alpha 7\beta 1$ integrin, the main integrin receptor of laminin-2 and laminin-4, remained unchanged in all genotypes (Table 1). In contrast, significant alterations were observed for the laminin $\alpha 4$ and $\alpha 5$ chains (Table 1 and Fig. 4a). In wild-type animals, both α -chains of laminin were mainly concentrated around blood vessels (Fig. 4a, $\alpha 4$ laminin and $\alpha 5$ laminin) and only a little immunoreactivity was associated with muscle basement membrane. In agreement with other studies^{9,10}, the amount of basement membrane-associated laminin $\alpha 4$ was strongly increased in *dy^w* mice, but was not further influenced by the mini-agrin (Fig. 4a, $\alpha 4$ laminin; see also Table 1). In contrast, the amount of the $\alpha 5$ chain was strongly influenced by mini-agrin (Fig. 4a, $\alpha 5$ laminin; see also Table 1). Notably, only a slight increase in the amount of the laminin $\alpha 5$ chain was observed in two mag-tg mice on wild-type background (animal 1, $123 \pm 5\%$; animal 2, $164 \pm 11\%$). Thus, mini-agrin-induced upregulation of the laminin $\alpha 5$ chain is also dependent on the genetic background of the mice.

Examination of α - and β -dystroglycan revealed that the two proteins were regulated differentially. This is notable as both proteins are derived from a single transcript and are generated by post-translational mechanisms²². The amount of the transmembrane component β -dystroglycan was approximately three times higher in both *dy^w* and *dy^w/mag-tg* mice (Fig. 4a, β -dystroglycan; see also Table 1). Despite the increase in β -dystroglycan, the amount of α -dystroglycan was decreased in *dy^w* mice (Fig. 4a, α -dystroglycan; see also Table 1). Expression of mini-agrin, which binds to α -dystroglycan, restored the amount of α -dystroglycan in *dy^w/mag-tg* mice to almost wild-type levels (Fig. 4a and Table 1). The change in the expression of α - and β -dystroglycan was again dependent on the genetic background as their levels in mag-tg mice were not different from wild-type mice (data not shown). The increase in the amount of α -dystroglycan and the laminin $\alpha 5$ chain in *dy^w/mag-tg* mice does not involve upregulation of gene transcription, as determined by polymerase chain reaction with reverse transcription (RT-PCR) (data not shown). Thus, the binding of mini-agrin to α -dystroglycan and $\alpha 5$ chain-containing laminins may stabilize them. To further investigate this we used Sepharose beads coupled to wheat-germ lectin (WGL), which bind to α -dystroglycan²³ but not to the mini-agrin protein (data not shown), to enrich α -dystroglycan from skeletal muscle extracts. As shown in Fig. 4c, the mini-agrin protein co-precipitated with WGL Sepharose, indicating that α -dystroglycan and the mini-agrin protein form a protein complex. The co-precipitation was specific as no enrichment of the mini-agrin protein was detected when muscle extracts from mag-tg mice were treated with non-conjugated Sepharose beads (data not shown).

Our findings demonstrate that an agrin mini-gene specifically

combining distinct binding properties is sufficient to amend muscle function in laminin $\alpha 2$ -deficient mice. In 16-week-old mice—an age at which most of the *dy^w* mice had died already—we observed signs of a mild myopathy in *dy^w/mag-tg* mice. This suggests that the mini-agrin may not fully rescue all of the muscle degeneration. The remaining pathology in the *dy^w/mag-tg* mice, such as hindlimb paralysis and cardiomyopathy, may also be caused by the lack of mini-agrin expression in the respective tissue, that is, heart and peripheral nerve. Our data indicate that amelioration of the myopathy in *dy^w/mag-tg* mice involves stabilization of α -dystroglycan. The observed restoration of muscle basement membrane structure in *dy^w/mag-tg* mice may also involve α -dystroglycan as α -dystroglycan deficiency abrogates proper basement membrane assembly²⁴ and can cause muscular dystrophy²⁵. The mechanism responsible for the mini-agrin-mediated rescue may also involve the laminin $\alpha 5$ chain. This α -chain includes the N-terminal domain VI, which is required for the formation of the laminin scaffold, and may thus contribute to the restoration of the muscle basement membrane structure. Like agrin, the expression of the laminin $\alpha 5$ chain is restricted to the neuromuscular junction in adult muscle⁹.

Our studies are an example of functional compensation of a diseased phenotype by an engineered molecule of the extracellular matrix, whose design was based on functional and not on structural similarity between the mutated laminin $\alpha 2$ chain and muscle agrin. Our results show that the use of small mimetics of large proteins can be useful for gene therapy. Such mimetics can be incorporated into existing viral vectors and are less likely to cause immunological responses. In the case of agrin, such gene therapy may not require high transfection efficiency because mini-agrin is secreted from muscle fibres and can act on neighbouring, non-transfected fibres. Furthermore, the use of pharmacologically active compounds that increase the level of expression of the endogenous agrin could be another promising strategy in the treatment of MCMD. □

Methods

Mini-agrin transgenic mice

The minigene derived from chick muscle agrin was generated analogously to the procedure described for a minigene derived from neural agrin²⁶. The 1.3-kb MCK promoter^{14,15} was inserted into the SK pBluescript vector (Stratagene) using a *HindIII* site. The construct was excised using *BssHII* and *XbaI* restriction enzymes and injected into mouse oocytes. Transgenic lines were identified by PCR and Southern blot analysis using primers and probes from the 3'-untranslated region of chick agrin mRNA. From eight transgenic founders, we used two lines that expressed the mini-agrin protein at high levels. We carried out genotyping of *dy^w* mice as described¹⁶.

Locomotion

We examined exploratory locomotion of mice in an open field test. In each experiment, one wild-type mouse and either a *dy^w* or a *dy^w/mag-tg* mouse was placed into a new cage and videotaped for 5 min. The time that mice spent moving around was measured manually. In the rotating rod task, animals were first trained three times to allow for accommodation to the task. After resting for at least 1 h, animals were tested with the rod turning at 16 r.p.m. Each mouse was tested three times for 2 min. When possible, the data were collected blindly.

Histology and quantification of fibre size

Muscles were immersed in 7% Gum Tragacanth (Sigma) and rapidly frozen in liquid-nitrogen-cooled isopentane. Sections (10- μ m thick) were stained with haematoxylin and eosin. For quantification, individual muscle sections were digitized and the number of squares (8 $\times$ 8 μ m) was counted per muscle fibre. For each genotype, four mice were used in which five sections were counted. Data were collected blindly.

Creatine kinase assay

Blood from 4-week-old mice was collected from the tail, and creatine kinase levels were measured in 2 μ l of serum using the creatine kinase 10 kit (Sigma). We included serum from wild-type mice for normalization.

Antibodies

We used the following antibodies: anti-chick agrin⁸, anti-nidogen (gift from M. Chiquet), anti-laminin $\alpha 4$ (ref. 11), anti-laminin $\alpha 5$ (ref. 27; R. Timpl, unpublished data), anti-laminin $\gamma 1$ (Chemicon), anti-perlecan²⁸, anti- $\alpha 7$ integrin⁵, anti- β -dystroglycan (Novo-Castra) and anti- α -dystroglycan²⁹.

Immunoblots and immunostaining

Muscles were rapidly homogenized in 1 ml extraction buffer (75 mM Tris-HCl, pH 6.8, 3.8% SDS, 4 M urea, 20% glycerol and 5% β -mercaptoethanol). After protein denaturation (95 °C for 5 min), non-dissolved protein was removed by centrifugation (14,000g for 5 min). Fifty micrograms of protein, as determined by Lowry protein assay, was loaded onto each lane, separated on 3–12% SDS–polyacrylamide gel electrophoresis (PAGE) and immunoblotted. For immunostaining, muscles were immediately immersed in optimal cutting temperature compound (OCT; Sakura Finetek) and frozen in liquid-nitrogen-cooled isopentane. Frozen sections (10- μ m thick) were incubated with blocking solution (2% horse serum, 1% BSA, 0.5% Triton X-100 in PBS) for 30 min. Primary antibodies were diluted in incubation buffer (2% horse serum, 1% BSA, 0.01% Triton X-100 in PBS) and incubated overnight at 4 °C. Sections were washed four times with incubation buffer and appropriate secondary antibodies were incubated at room temperature for 2 h. To determine protein expression levels, the primary antibodies were diluted such that the signal was not saturated. Images were collected and analysed with a confocal microscope (Leica TCS-8P) and appropriate software. The microscope was calibrated using the InSpeck Microscope Image Intensity Calibration Kit (Molecular Probes). Specific intensity was calculated for each image as the signal intensity of the muscle circumference minus that of an adjacent, non-stained region³⁰. Five different pictures were taken using the same parameters on each section, and four different sections were used for each individual mouse.

Electron microscopy

Mice were perfused with 2.5% glutaraldehyde in 0.1 M sodium phosphate, pH 7.4. Muscles were post-fixed overnight at 4 °C in 2.5% glutaraldehyde in 0.1 M sodium phosphate. After treatment with 1% OsO₄ in 0.1 M cacodylate buffer (pH 7.2) muscle pieces were rinsed in 1% Na₂SO₄ in 0.1 M cacodylate buffer and imbedded in Epon. Sections (60-nm thick) were cut on a microtome (Ultracut E; Leica) and post-stained with uranyl acetate and lead citrate. The specimens were examined with an electron microscope (Philips EM400) at an accelerating voltage of 80 kV. We took pictures without knowing the genotype of the mice.

Lectin precipitation

Muscles were homogenized in 500 μ l incubation buffer (50 mM Tris-HCl, pH 7.5, 0.2 M NaCl, 1.25 mM CaCl₂, 1 mM MgCl₂). After centrifugation at 20,000g, the supernatant was collected and the pellet was re-extracted with 1% Triton X-100 in incubation buffer for 30 min on ice. After centrifugation at 20,000g, the supernatant was collected and pooled with the supernatant from the first centrifugation. One hundred microlitres of wheat germ lectin (WGL) Sepharose beads (Amersham Pharmacia) was added and incubated for 2 h at 4 °C. WGL beads were washed four times with incubation buffer and proteins were eluted in 75 mM Tris-HCl, pH 6.8, 3.8% SDS, 4 M urea, 20% glycerol and 5% β -mercaptoethanol. Proteins were separated on 3–12% SDS–PAGE and immunoblotted using anti-chick agrin antibodies.

Received 5 March; accepted 23 July 2001.

1. Tome, F. M. S., Guicheney, P. & Fardeau, M. in *Neuromuscular Disorders: Clinical and Molecular Genetics* (ed. Emery, A. E. H.) 21–57 (John Wiley & Sons, West Sussex, 1998).
2. Miyagoe-Suzuki, Y., Nakagawa, M. & Takeda, S. Merosin and congenital muscular dystrophy. *Microsc. Res. Tech.* **48**, 181–191 (2000).
3. Colognato, H. & Yurchenco, P. D. Form and function: the laminin family of heterotrimers. *Dev. Dyn.* **218**, 213–234 (2000).
4. Campbell, K. P. Three muscular dystrophies: loss of cytoskeleton-extracellular matrix linkage. *Cell* **80**, 675–679 (1995).
5. Mayer, U. *et al.* Absence of integrin $\alpha 7$ causes a novel form of muscular dystrophy. *Nature Genet.* **17**, 318–323 (1997).
6. Ruegg, M. A. & Bixby, J. L. Agrin orchestrates synaptic differentiation at the vertebrate neuromuscular junction. *Trends Neurosci.* **21**, 22–27 (1998).
7. Denzer, A. J., Brandenberger, R., Gesemann, M., Chiquet, M. & Ruegg, M. A. Agrin binds to the nerve-muscle basal lamina via laminin. *J. Cell Biol.* **137**, 671–683 (1997).
8. Gesemann, M., Brancaccio, A., Schumacher, B. & Ruegg, M. A. Agrin is a high-affinity binding protein of dystroglycan in non-muscle tissue. *J. Biol. Chem.* **273**, 600–605 (1998).
9. Patton, B. L., Miner, J. H., Chiu, A. Y. & Sanes, J. R. Distribution and function of laminins in the neuromuscular system of developing, adult, and mutant mice. *J. Cell Biol.* **139**, 1507–1521 (1997).
10. Ringelmann, B. *et al.* Expression of laminin $\alpha 1$, $\alpha 2$, $\alpha 4$, and $\alpha 5$ chains, fibronectin, and tenascin-C in skeletal muscle of dystrophic 129ReJ dy/dy mice. *Exp. Cell Res.* **246**, 165–182 (1999).
11. Talts, J. F. *et al.* Structural and functional analysis of the recombinant G domain of the laminin $\alpha 4$ chain and its proteolytic processing in tissues. *J. Biol. Chem.* **275**, 35192–35199 (2000).
12. Denzer, A. J. *et al.* Electron microscopic structure of agrin and mapping of its binding site in laminin-1. *EMBO J.* **17**, 335–343 (1998).
13. Burgess, R. W., Nguyen, Q. T., Son, Y. J., Lichtman, J. W. & Sanes, J. R. Alternatively spliced isoforms of nerve- and muscle-derived agrin: their roles at the neuromuscular junction. *Neuron* **23**, 33–44 (1999).
14. Jaynes, J. B., Chamberlain, J. S., Buskin, J. N., Johnson, J. E. & Hauschka, S. D. Transcriptional regulation of the muscle creatine kinase gene and regulated expression in transfected mouse myoblasts. *Mol. Cell. Biol.* **6**, 2855–2864 (1986).
15. Sternberg, E. A. *et al.* Identification of upstream and intragenic regulatory elements that confer cell-type-restricted and differentiation-specific expression on the muscle creatine kinase gene. *Mol. Cell. Biol.* **8**, 2896–2909 (1988).
16. Kuang, W. *et al.* Merosin-deficient congenital muscular dystrophy. Partial genetic correction in two mouse models. *J. Clin. Invest.* **102**, 844–852 (1998).
17. Sunada, Y., Bernier, S. M., Kozak, C. A., Yamada, Y. & Campbell, K. P. Deficiency of merosin in dystrophic dy mice and genetic linkage of laminin M chain gene to dy locus. *J. Biol. Chem.* **269**, 13729–13732 (1994).

18. Xu, H., Wu, X. R., Wewer, U. M. & Engvall, E. Murine muscular dystrophy caused by a mutation in the laminin $\alpha 2$ (Lama2) gene. *Nature Genet.* **8**, 297–302 (1994).
19. Xu, H., Christmas, P., Wu, X. R., Wewer, U. M. & Engvall, E. Defective muscle basement membrane and lack of M-laminin in the dystrophic dy/dy mouse. *Proc. Natl Acad. Sci. USA* **91**, 5572–5576 (1994).
20. Colognato, H. & Yurchenco, P. D. The laminin $\alpha 2$ expressed by dystrophic dy(2J) mice is defective in its ability to form polymers. *Curr. Biol.* **9**, 1327–1330 (1999).
21. MacPike, A. D. & Meier, H. Comparison of dy and dy2J, two alleles expressing forms of muscular dystrophy in the mouse. *Proc. Soc. Exp. Biol. Med.* **151**, 670–672 (1976).
22. Ibraghimov-Beskrovnya, O. *et al.* Primary structure of dystrophin-associated glycoproteins linking dystrophin to the extracellular matrix. *Nature* **355**, 696–702 (1992).
23. Ervasti, J. M., Kahl, S. D. & Campbell, K. P. Purification of dystrophin from skeletal muscle. *J. Biol. Chem.* **266**, 9161–9165 (1991).
24. Henry, M. D. & Campbell, K. P. A role for dystroglycan in basement membrane assembly. *Cell* **95**, 859–870 (1998).
25. Cote, P. D., Mouchles, H., Lindenbaum, M. & Carbonetto, S. Chimaeric mice deficient in dystroglycans develop muscular dystrophy and have disrupted myoneural synapses. *Nature Genet.* **23**, 338–342 (1999).
26. Meier, T. *et al.* A minigene of neural agrin encoding the laminin-binding and acetylcholine receptor-aggregating domains is sufficient to induce postsynaptic differentiation in muscle fibres. *Eur. J. Neurosci.* **10**, 3141–3152 (1998).
27. Sorokin, L. M., Pausch, F., Durbeej, M. & Ekblom, P. Differential expression of five laminin α (1–5) chains in developing and adult mouse kidney. *Dev. Dyn.* **210**, 446–462 (1997).
28. Schulze, B., Mann, K., Battistutta, R., Wiedemann, H. & Timpl, R. Structural properties of recombinant domain III-3 of perlecan containing a globular domain inserted into an epidermal-growth-factor-like motif. *Eur. J. Biochem.* **231**, 551–556 (1995).
29. Herrmann, R. *et al.* Dissociation of the dystroglycan complex in caveolin-3-deficient limb girdle muscular dystrophy. *Hum. Mol. Genet.* **9**, 2335–2340 (2000).
30. Turney, S. G., Culican, S. M. & Lichtman, J. W. A quantitative fluorescence-imaging technique for studying acetylcholine receptor turnover at neuromuscular junctions in living animals. *J. Neurosci. Methods* **64**, 199–208 (1996).

Acknowledgements

We are grateful to M. Chiquet, S. Kröger, U. Mayer, L. Sorokin and R. Timpl for providing us with antibodies. We thank S. Arber, A. Brancaccio and T. Meier for discussions. M. Dürrenberger and U. Sauder for help in electron microscopy; J.-F. Spetz for DNA injection; M. Willem for providing us with the MCK promoter; and P. Scotton for the help with artwork. Special thanks go to D. Walz for his help in the statistical analysis of the results. This work was supported by grants from the Swiss National Science Foundation (MAR), the Kanton of Basel-Stadt (MAR), the Swiss Foundation for Research on Muscle Diseases (MAR and UM), from the National Institutes of Health (EE), the Novartis Research Foundation (UM), the Deutsche Forschungsgemeinschaft, Aktion Benni and company, and the Ernst und Berta Grimmke-Stiftung (HL).

Correspondence and requests for materials should be addressed to M.A.R. (e-mail: markus-a.ruegg@unibas.ch).

Intercellular movement of the putative transcription factor SHR in root patterning

Keiji Nakajima*, Giovanni Sena, Tal Nawy & Philip N. Benfey

Department of Biology, New York University, 1009 Main Building, 100 Washington Square East, New York, New York 10003, USA

Positional information is pivotal for establishing developmental patterning in plants^{1–3}, but little is known about the underlying signalling mechanisms. The *Arabidopsis* root radial pattern is generated through stereotyped division of initial cells and the subsequent acquisition of cell fate⁴. *short-root* (*shr*) mutants do not undergo the longitudinal cell division of the cortex/endodermis initial daughter cell, resulting in a single cell layer with only cortex attributes^{5,6}. Thus, *SHR* is necessary for both cell division and endodermis specification^{5,6}. *SHR* messenger RNA is found exclusively in the stele cells internal to the endodermis and cortex, indicating that it has a non-cell-autonomous mode of action⁶.

* Present address: Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0101, Japan.

Here we show that the SHR protein, a putative transcription factor, moves from the stele to a single layer of adjacent cells, where it enters the nucleus. Ectopic expression of *SHR* driven by the promoter of the downstream gene *SCARECROW* (*SCR*) results in autocatalytic reinforcement of SHR signalling, producing altered cell fates and multiplication of cell layers. These results support a model in which SHR protein acts both as a signal from the stele and as an activator of endodermal cell fate and SCR-mediated cell division.

To determine whether SHR protein could move between cells, we fused in-frame the coding regions of SHR and green fluorescent protein (GFP) under the control of the 2.5-kilobase (kb) 5'-upstream region of *SHR*. This promoter element confers an expression pattern that matches the localization of *SHR* mRNA⁶ (Fig. 1b, insert). Transforming the *shr-2* mutant (a presumed null allele⁶) with the fusion construct resulted in rescue of the root radial pattern (Fig. 1b) with appropriate cell-fate specification (data not shown), indicating that the fusion protein retained the SHR activity necessary for root patterning.

We analysed the localization of SHR::GFP fluorescence by both

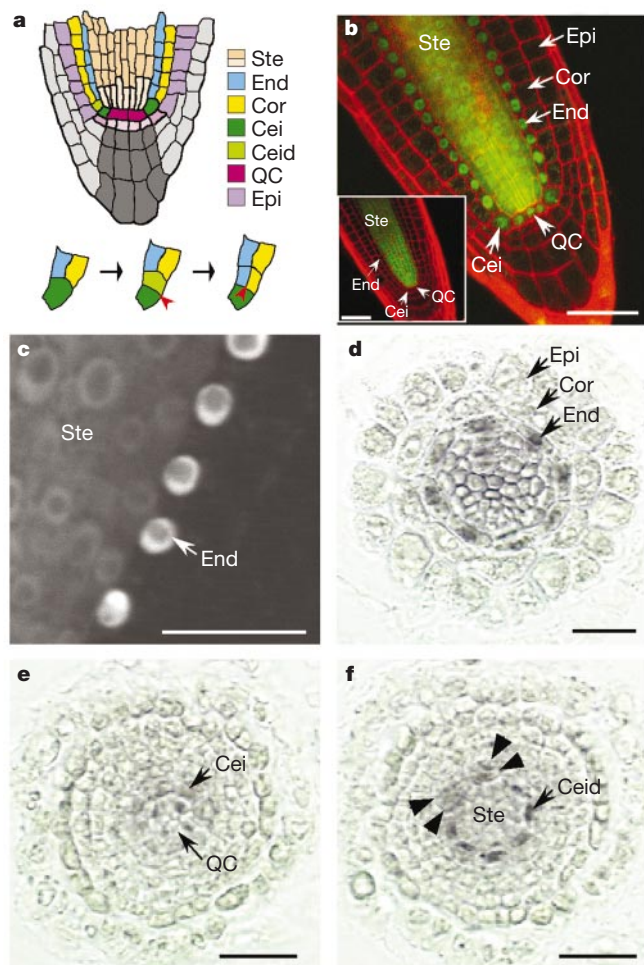


Figure 1 SHR protein localization. **a**, Diagram of *Arabidopsis* root tip and cell divisions (red arrowheads) that form endodermis and cortex. **b**, **c**, Confocal (**b**) and multiphoton (**c**) images of the SHR::GFP transgenic roots (insert in **b**, transcriptional fusion⁶). **d–f**, SHR immunostaining on transverse root sections above root meristem (**d**), at the level of the QC (**e**) and Ceid after longitudinal division (**f**). Arrowheads in **f** indicate SHR protein in two Ceid daughters. Cei, cortex/endodermis initial; Ceid, daughter of Cei; Cor, cortex; End, endodermis; Epi, epidermis; Ste, stele; QC, quiescent centre. Scale bars, 50 μ m (**b**), 25 μ m (**c–f**).

confocal and multi-photon laser-scanning microscopy. GFP fluorescence was found not only in the stele, but also in all cells immediately adjacent to the stele (Fig. 1b). These include the quiescent centre (QC), the cortex/endodermis initial and daughter cells, as well as all cells of the endodermis (Fig. 1a) (we refer to these cells collectively as the 'adjacent layer'). *SHR* transcripts have never been detected in any of these cells either by *in situ* hybridization⁶ or indirectly through *SHR* promoter reporter gene fusions⁶ (Fig. 1b, insert). From the confocal images, it appeared that although GFP fluorescence was localized to both the nuclei and the cytoplasm in the stele, it was primarily localized to the nuclei in the adjacent layer (Fig. 1b). This was confirmed with multi-photon microscopy (Fig. 1c). At the limit of resolution of multi-photon imaging, we could not detect GFP in cortex or epidermal cells, nor in any other organelles or subcellular structures besides the nuclei in endodermal cells (Fig. 1c).

To confirm this localization of SHR protein, we produced affinity-purified antibodies to recombinant SHR protein. The specificity of the antibody was analysed on *shr-2* mutant roots where no binding was detected (Supplementary Information Fig. 1b). On transverse sections from just above the root meristem, the antibodies consistently decorated nuclei of endodermal cells (Fig. 1d). In the stele, there was more uniform binding of the antibodies (Fig. 1d), with occasional stronger staining observed in nuclei (Supplementary Information Fig. 1a). In sections at the level of the QC, nuclear localized SHR protein was evident in the four QC cells, as well as in the eight surrounding cortex/endodermis initial cells (Fig. 1e). In the next serial section, strong staining was detected in the initial daughter cells after their longitudinal cell division (Fig. 1f).

Notably, in two of the eight pairs of daughter cells SHR was localized in both nuclei resulting from the longitudinal division (Fig. 1f, triangles), whereas in the remaining six pairs only the inner nucleus appears to contain SHR protein (Fig. 1f). This is probably due to the non-synchronous division of these cells, which results in slightly different times after division being analysed in the pairs of nuclei. Together, these results indicate that SHR is equally partitioned into the two daughter nuclei after the longitudinal cell division, but SHR levels are maintained only in the nucleus of the inner cell.

The simplest explanation for the difference in localization patterns of *SHR* mRNA and protein is movement of SHR protein from the stele to the adjacent layer. An alternative hypothesis is an abnormally rapid degradation of *SHR* mRNA transcribed in or transported to the adjacent layer. We have shown, however, that when *SHR* is driven by the constitutive cauliflower mosaic virus 35S (CaMV35S) promoter, high levels of *SHR* mRNA accumulate in cells outside the stele⁶. Moreover, both the rapid loss of SHR immunostaining in outer daughter cells of the longitudinal division (Fig. 1f) and the lack of GFP fluorescence in the first cell of the cortex lineage in the SHR::GFP lines (Fig. 1b) indicate that SHR protein has a relatively short half-life. Thus, the discrepancy between the mRNA and protein localization is not due to the protein having a much lower turnover rate than the mRNA. We therefore conclude that SHR protein moves from the stele to the single adjacent layer.

To understand better the role of SHR protein in specifying cell fate and mediating cell division in the root, we generated transgenic plants in which *SHR* is transcribed in the adjacent layer. The 2.5-kb *SCR* promoter, which confers expression in all cells of the adjacent layer^{6,7}, was fused to the full-length *SHR* coding region (*SCRpro::SHR*). This construct was introduced into wild-type plants either with or without a *SCRpro::GFP* reporter. Thirteen independent lines showed an increased number of cell layers in the root (Fig. 2a, b). The supernumerary cell layers exhibited a high degree of radial symmetry. This is in contrast to a phenotype obtained with ectopic *SHR* expression by the CaMV35S promoter,

which displayed fairly chaotic division patterns⁶.

The differentiation status of root cells in the *SCRpro::SHR* lines was initially analysed with two monoclonal antibodies specific to epitopes found in the different cell layers: the JIM13 antibody recognizes an arabinogalactan epitope on the endodermis and on some stele cells⁸ (Supplementary Information Fig. 2b), and the CCRC-M2 antibody recognizes a carbohydrate epitope in cortex and epidermis cell walls⁹ (Supplementary Information Fig. 2c). JIM13 stained all the cell layers except the outermost layer, indicating the presence of endodermal characters in these cells (Fig. 2c). In the outermost layer a few cells were occasionally stained with JIM13. When the same section was treated with CCRC-M2, only the outermost layer was stained (Fig. 2d). Notably, the JIM13-stained cells in the outermost layer were not stained by the CCRC-M2 antibody (Fig. 2c and d, asterisks), suggesting that endodermal specification can override cortex and/or epidermal specification. Acquisition of endodermal characters in the supernumerary layers was confirmed by histochemical staining for the Casparian strip, a cell wall deposit of hydrophobic carbohydrate that occurs specifically in the endodermis¹⁰ (Supplementary Information Fig. 2a). Staining of the Casparian strip was detected in most, if not all, of the cells between the stele and the outermost cell layer (Fig. 2e).

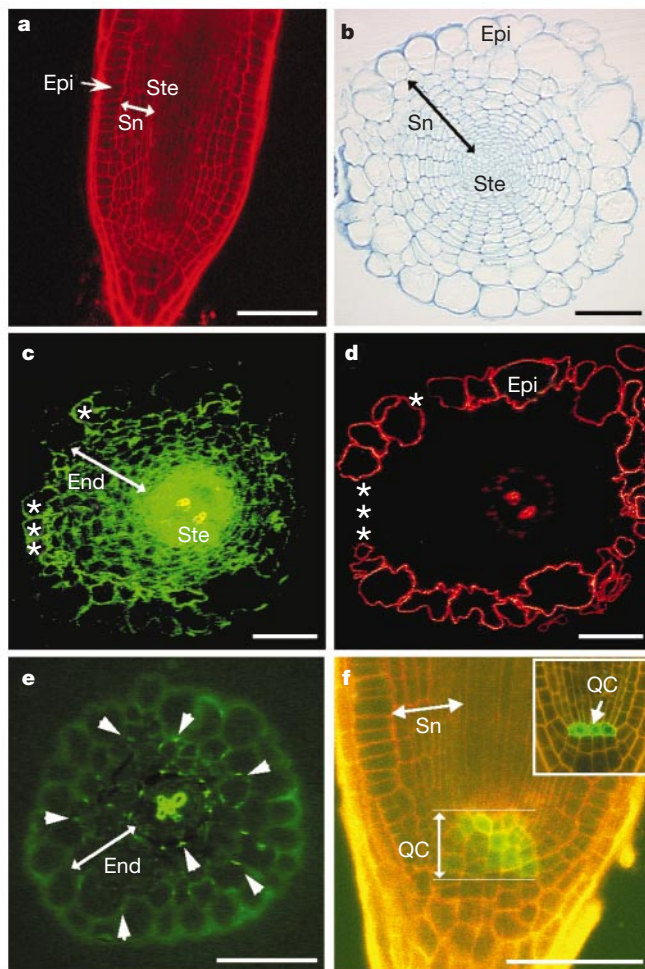


Figure 2 Analysis of *SCRpro::SHR* transgenic roots. **a, b**, Confocal image (**a**) and transverse section (**b**) of the *SCRpro::SHR* transgenic root. **c–f**, Marker analysis of the *SCRpro::SHR* root. **c, d**, JIM13 (**c**) and CCRC-M2 (**d**) antibody stains of the same section. Asterisks indicate the same cells in **c** and **d**. **e**, Casparian strip stain. **f**, Confocal images of the QC433 QC-specific GFP marker (inset, wild type). Sn, supernumerary layers. Other abbreviations as in Fig. 1. Scale bars, 50 μm.

Because SHR protein seems to move not only to the endodermis but also to the QC, we determined the status of QC cells in the *SCRpro::SHR* lines. A GFP marker line, QC433, that exhibits highly specific expression in the QC (Fig. 2f, insert) was crossed to a strong *SCRpro::SHR* line. The F₁ plants had an increased number of GFP-expressing cells as compared with wild-type roots (Fig. 2f). These cells were organized into supernumerary layers between the stele and the root cap. The supernumerary endodermis layers appear to converge on these GFP-expressing cells (Fig. 2f). Crosses to root-cap-specific marker lines indicated that neither the lateral root cap nor the columella root cap was affected in the transgenic roots (data not shown).

To determine whether the cell-fate changes observed in the *SCRpro::SHR* lines correlated with the presence of SHR protein we performed immunolocalization with the anti-SHR antibodies in backgrounds that contained tissue-specific reporters. We crossed a strong *SCRpro::SHR* line with the marker lines, End195 (ref. 11), *GL2pro::GUS* (ref. 12) and QC46 (ref. 13), which show specific β-glucuronidase (GUS) expression in endodermis, non-hair epidermal cells and QC cells, respectively. The roots of plants from each cross were first stained for GUS activity, and then fixed, sectioned and processed for SHR immunolocalization. Consistent with the

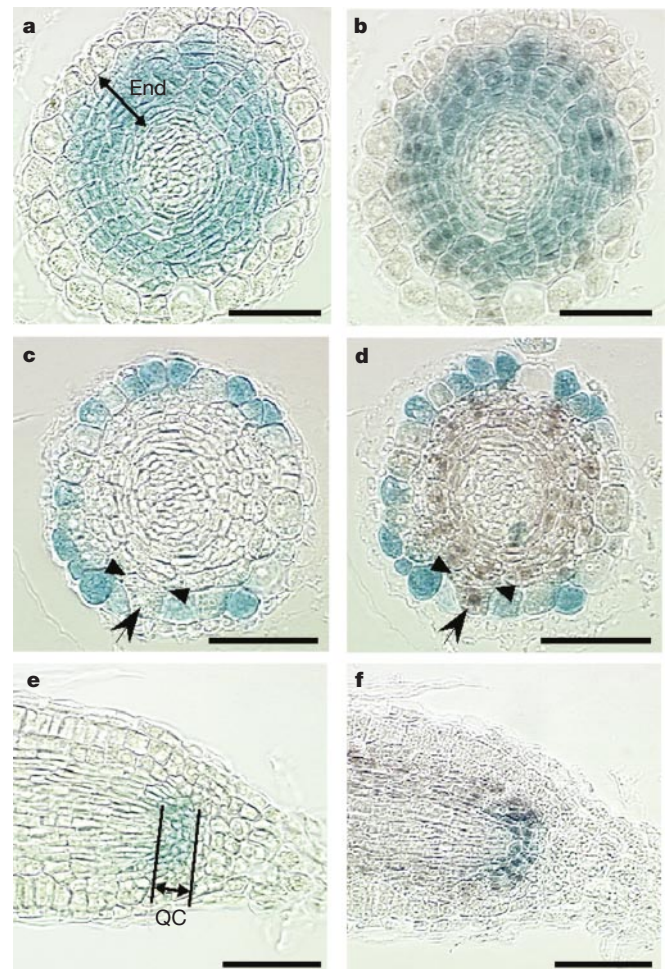


Figure 3 Correlation between the presence of SHR protein and cell fates in *SCRpro::SHR* roots. GUS-stained sections of the crosses between *SCRpro::SHR* and End195 (**a** and **b**), *GL2pro::GUS* (**c** and **d**), or QC46 (**e** and **f**), either before (**a, c** and **e**) or after (**b, d** and **f**) SHR immunostaining. Arrows in **c** and **d** indicate presence of SHR protein in a cell of the outermost layer that normally should express *GL2*, and triangles point to both ends of an underlying cell. Abbreviations as in Fig. 1. Scale bars, 50 μm.

supernumerary cells having endodermal characters, the End195 GUS stain was found in all cells between the stele and the outermost layer (Fig. 3a). Immunolocalization of SHR on the same section showed a strong signal in the GUS-expressing cells, whereas staining was rarely detected in cells of the outermost layer (Fig. 3b).

In a similar analysis of the cross with the *GL2pro::GUS* line, GUS staining was detected only in cells of the outermost layer that were in contact with a single underlying cell (Fig. 3c). Contact with two underlying cortex cells is normally required for the acquisition of hair cell fate, and *GL2* is specifically expressed in the non-hair epidermal cells¹². *GL2* expression in the *SCRpro::SHR* background is therefore analogous to wild type, indicating that the underlying cells need not be cortex for correct specification to occur. Occasionally we found cells in the outermost layer in which there was no GUS staining even though they were situated over a single underlying cell (Fig. 3c, arrow). The SHR immunostain revealed the presence of SHR protein in these cells (Fig. 3d, arrow). These results indicate that when SHR is found in cells of the outermost layer it is sufficient to alter their cell fate.

In the analysis of the cross with the QC46 line, the GUS stain was found in the 3–4 cell layers that extend toward the root cap from the base of the vascular cylinder (Fig. 3e). Immunostaining of this section revealed the presence of the SHR protein in all of the GUS-expressing layers (Fig. 3f), suggesting that production of multiple QC layers was a result of SHR in these cells. The formation of multiple QC layers should give rise to multiple layers of endodermis/cortex initial cells immediately adjacent to them. This idea is based on laser ablation experiments that showed that QC cells have a role in maintaining the surrounding initial cells by repressing their differentiation². Formation of the supernumerary endodermis layers is unlikely to result solely from the presence of the ectopic QC layers, however, because the number of supernumerary endodermis layers exceeds by more than twofold the number of QC layers.

To determine whether *SCR* is involved in production of the supernumerary layers, we analysed *SCR* transcription in the *SCRpro::SHR* lines. Previous analyses showed that levels of *SCR* RNA were sharply reduced in *shr* mutants, suggesting that *SHR* is required for full transcriptional activation of *SCR*⁶. Consistent with this model, we observed strong GFP fluorescence from the *SCRpro::GFP* marker in the supernumerary cell layers (Fig. 4a). The presence of SHR protein correlated well with the GFP expression, as revealed by immunostaining of serial sections with either SHR or GFP antibodies (data not shown).

To determine whether *SCR* activity was necessary for the ectopic cell divisions in the *SCRpro::SHR* lines, we crossed a strong *SCRpro::SHR* line with the *scr-1* mutant, which has no detectable *SCR* RNA⁷. In plants carrying the *SCRpro::SHR* transgene and homozygous for the *scr-1* allele, the radial pattern was identical to that of the *scr* mutant (Fig. 4b). The progeny of a backcross to wild type had the *SCRpro::SHR* radial pattern, whereas all selfed F₃

progeny tested had the *scr* mutant radial pattern (data not shown). This shows that the lack of ectopic cell divisions in the *scr* background was not a consequence of inactivation of the transgene *per se*. In conclusion, the *SCRpro::SHR* phenotype requires a functional *SCR* gene, indicating that SHR activation of *SCR* is necessary for the longitudinal cell divisions that are essential for radial patterning.

Although the longitudinal division of the cortex/endodermal initial daughter does not take place in *scr* mutants, the resulting single mutant cell layer has differentiated attributes of both cortex and endodermis⁷. To determine whether SHR protein can still translocate to the single mutant tissue layer, we crossed the SHR::GFP fusion protein line to the *scr-1* mutant. In plants carrying the SHR::GFP transgene and homozygous for the *scr-1* allele, GFP fluorescence was clearly visible in the nuclei of the mutant layer (Fig. 4c). We conclude that SHR protein movement and nuclear localization do not require *SCR* activity.

Movement of SHR protein provides a simple model to explain the non-cell-autonomous function of *SHR* required for root radial patterning. In this hypothesis, SHR has three functions: transmission of positional information through its own movement; activation of *SCR*; and cell-fate specification (Supplementary Information Fig. 3). The last two functions would require movement of SHR protein. To test this model, we enlarged the SHR-expression domain and analysed how well cell differentiation and activation of *SCR* correlated with localization of SHR protein. All the observed phenotypes can be explained by applying the model. In transgenic *SCRpro::SHR* plants, activation of *SCR* by SHR results in *in situ* production of SHR in the adjacent layer, because the SHR coding region is under control of the *SCR* promoter. This creates an autocatalytic reinforcement process of SHR signalling. The consequences of this reinforcement are *SCR*-dependent cell division and replacement of cortex (and occasionally epidermis) cell fate with endodermis cell fate, which is mediated by the cell-fate specification function of SHR. The ectopic presence of SHR in the outer daughter cell could be due to influx of the SHR produced from the *SCRpro::SHR* transgene in the inner daughter cell, or to *de novo* transcription from the transgene in the outer daughter cell mediated by the residual SHR, or to a combination of both. In the cells immediately below the stele, the reinforcement process and cell-fate specification result in multiple layers of QC. This indicates that cell-fate specification by SHR requires yet unidentified positional cues that are not mediated by SHR, because the presence of SHR can result in different fates when cells are in different positions.

Although our study does not address the pathway of SHR movement, previous work suggests that movement is likely to be through plasmodesmata¹⁴. In the shoot meristem, intercellular trafficking of transcription factors has been reported for structurally unrelated proteins^{15,16}. The biological relevance of this movement is not evident because in wild-type plants these proteins are found in an identical domain as their cognate RNA. The one exception, the maize KNOTTED1 (KN1) homeobox protein has been reported to move to the outermost layer of the shoot meristem where its RNA was not detected¹⁷. But there is no evidence that protein presence in the outermost layer is essential for KN1 activity. In contrast, because the non-cell-autonomous activity of SHR is essential for patterning the adjacent cell layer, movement of SHR protein seems to have an integral role in root radial pattern formation.

In summary, we have shown that SHR protein can move and its translocation provides a simple model for its non-cell-autonomous activity in root radial patterning. Nuclear localization of SHR in the adjacent layer and activation of *SCR* strongly suggest that SHR is directly involved in transcriptional regulation in these cells. This simple strategy, which uses a single protein for both positional signalling and cell-fate determination, may have been important in the evolution of land plants to acquire efficient vascular function by forming radially organized tissues. □

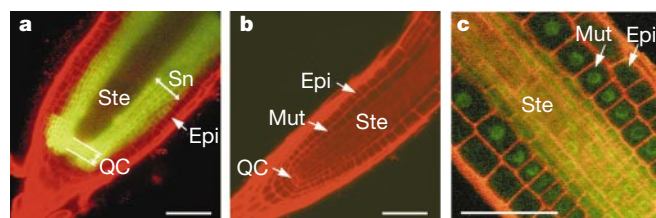


Figure 4 *SCR* acts downstream of *SHR*. **a**, *SCRpro::GFP* marker expression in a *SCRpro::SHR* transgenic root. **b**, Root phenotype of a *SCRpro::SHR* transgenic in the homozygous *scr-1* mutant background is the same as the *scr-1* phenotype. **c**, Expression of the SHR::GFP transgene in the homozygous *scr-1* mutant background results in detectable GFP fluorescence in the nuclei of the mutant layer. Mut, mutant layer. Other abbreviations as in Figs 1 and 2. Scale bars, 50 μ m.

Methods

Plasmid construction and plant transformation

To generate an in-frame fusion of SHR and GFP, we introduced restriction sites at both ends of the coding sequences for SHR and mGFP4 (with an S65T mutation¹⁸). A linker sequence encoding Asp-Pro-Gly was inserted between SHR and GFP. The resulting DNA fragment was used to replace the coding region of a 4.1-kb *SHR* gene fragment (including the 2.5-kb 5'-upstream region) fused to the nopaline synthase 3' region in the binary vector pBIN19. To make the *SCRpro::SHR* construct, we fused a 2.5-kb *SCR* 5' region and the 1.6-kb *SHR* coding region in the pBIH vector, which had been modified from pBIN19 to confer hygromycin resistance. The plasmids were electroporated into *Agrobacterium* strain GV3-101, which then was used to transform *Arabidopsis* plants by the floral dip method¹⁹.

Marker analysis and microscopy

The Casparian strip, and JIM13 and CCRC-M2 antibody stains were performed as described⁶. We carried out JIM13 and CCRC-M2 antibody stains on the same section using appropriate secondary antibodies (rat and mouse, respectively). GFP fluorescence and propidium-iodide-stained cell walls in the root were observed by confocal microscopy as described²⁰. Multi-photon microscopy was performed on a Bio-Rad MRC1024 MP multiphoton microscope equipped with a Coherent Mira 900-F infrared pulsed laser with operation wavelength of 935 nm. We carried out GUS staining as described¹¹. Some microscopic images were processed with Photoshop (Adobe Systems) to increase contrast.

Antibody preparation

The SHR polypeptide lacking the amino-terminal 142 amino acids was expressed as a fusion protein to either thioredoxin (TRX::SHR) or maltose-binding protein (MBP::SHR) using pET32b (Novagen) and pET-MAL, respectively. pET-MAL was a hybrid plasmid made from pET40b (Novagen) and pMALc2x (New England Biolabs). We used TRX::SHR to raise antisera, and immobilized MBP::SHR on a gel matrix for affinity purification of the antisera. The proteins were induced in *Escherichia coli* strain BL21 (DE3). The insoluble TRX::SHR protein was solubilized with 8 M urea and purified with the HisBind Resin & Buffer Kit (Novagen). The purified TRX::SHR was dialysed, concentrated and injected into rabbits at Covance Research Products. The soluble MBP::SHR protein was affinity purified with Amylose Resin (New England Biolabs) and then coupled to NHS-activated Sepharose 4 Fast Flow (Amersham Pharmacia). In parallel, we extracted total soluble protein from the *shr-2* mutant roots and immobilized it on the Sepharose gel as above.

The antisera were diluted 10-fold in TBST buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% (v/v) Tween 20), and incubated with the MBP::SHR gel overnight with gentle agitation at 4 °C. The gel was washed once with TBST and once with CBST (50 mM sodium citrate pH 4.0, 150 mM NaCl, 0.05% (v/v) Tween 20). The bound antibodies were then eluted with GBST (50 mM glycine-HCl pH 2.5, 150 mM NaCl, 0.05% (v/v) Tween 20), and immediately neutralized by adding a 1/10 volume of 1 M Tris-HCl pH 7.5. This fraction was further absorbed to the *shr-2* protein gel for 7 h at 4 °C. The final unbound fraction was used for immunolocalization. Pre-immune serum purified in the same way showed no specific binding to wild-type root sections (data not shown).

Immunohistochemistry

Arabidopsis roots were fixed in 4% (w/v) paraformaldehyde in PBS (10 mM sodium phosphate pH 7.5, 130 mM NaCl) overnight at 4 °C, washed twice with PBS and embedded in 1% (w/v) agarose. The roots in excised agarose blocks were passed through an ethanol series followed by HemoDe (Fisher), and then embedded in Paraplast X-tra (Fisher). Sections (9 µm) were cut and placed on silane-coated slides (Polysciences). After dewaxing and rehydration, the sections were treated with HCl and proteinase K as described¹⁶ with minor modifications (5 min in 0.2 N HCl and 10 min in 250 µg ml⁻¹ proteinase K). After blocking with TBST containing 1% (w/v) bovine serum albumin and 1% (v/v) goat serum for 4 h at room temperature, the sections were incubated with the purified anti-SHR antibodies (1/1,200 dilution relative to the antiserum) in the blocking solution overnight at room temperature. Slides were given six 10-min washes with TBST and incubated with alkali-phosphatase-conjugated anti-rabbit immunoglobulin-γ (1/1,000 dilution, Vector Lab) for 2 h at room temperature. After washing four times in TBST and twice in TBS (TBST lacking Tween 20), the signal was developed with NBT/BCIP (Roche Molecular Biochemicals) in 0.1 M Tris-HCl pH 9.5, 0.1 M NaCl, 1 mM levamisole for 2 h at room temperature.

Received 18 May; accepted 1 August 2001.

- van den Berg, C., Willemsen, V., Hage, W., Weisbeek, P. & Scheres, B. Cell fate in the *Arabidopsis* root meristem determined by directional signalling. *Nature* **378**, 62–65 (1995).
- van den Berg, C., Willemsen, V., Hendriks, G., Weisbeek, P. & Scheres, B. Short-range control of cell differentiation in the *Arabidopsis* root meristem. *Nature* **390**, 287–289 (1997).
- Kidner, C., Sundaresan, V., Roberts, K. & Dolan, L. Clonal analysis of the *Arabidopsis* root confirms that position, not lineage, determines cell fate. *Planta* **211**, 191–199 (2000).
- Dolan, L. *et al.* Cellular organisation of the *Arabidopsis thaliana* root. *Development* **119**, 71–84 (1993).
- Benfey, P. N. *et al.* Root development in *Arabidopsis*: Four mutants with dramatically altered root morphogenesis. *Development* **119**, 57–70 (1993).
- Helariutta, Y. *et al.* The *SHORT-ROOT* gene controls radial patterning of the *Arabidopsis* root through radial signaling. *Cell* **101**, 555–567 (2000).
- Di Laurenzio, L. *et al.* The *SCARECROW* gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. *Cell* **86**, 423–433 (1996).

- Knox, J. P., Linstead, P. J., King, J., Cooper, C. & Roberts, K. Pectin esterification is spatially regulated both within cell walls and between developing tissues of root apices. *Planta* **181**, 512–521 (1990).
- Freshour, G. *et al.* Developmental and tissue-specific structural alterations of the cell wall polysaccharides of *Arabidopsis thaliana* roots. *Plant Physiology* **110**, 1413–1429 (1996).
- Esau, K. *Anatomy of Seed Plants* 215–242 (Wiley, New York, 1977).
- Malamy, J. & Benfey, P. Organization and cell differentiation in lateral roots of *Arabidopsis thaliana*. *Development* **124**, 33–44 (1997).
- Masucci, J. D. *et al.* The homeobox gene *GLABRA2* is required for position-dependent cell differentiation in the root epidermis of *Arabidopsis thaliana*. *Development* **122**, 1253–1260 (1996).
- Sabatini, S. *et al.* An auxin-dependent distal organizer of pattern and polarity in the *Arabidopsis* root. *Cell* **99**, 463–472 (1999).
- Lucas, W. *et al.* Selective trafficking of KNOTTED1 homeodomain protein and its mRNA through plasmodesmata. *Science* **270**, 1980–1983 (1995).
- Perbal, M., Haughn, G., Siedler, H. & Schwarz-Sommer, Z. Non-cell-autonomous function of the Antirrhinum floral homeotic proteins *DEFICIENS* and *GLOBOSA* is exerted by their polar cell-to-cell trafficking. *Development* **122**, 3433–3441 (1996).
- Sessions, A., Yanofsky, M. F. & Weigel, D. Cell–cell signaling and movement by the floral transcription factors *LEAFY* and *APETALA1*. *Science* **287**, 419–421 (2000).
- Jackson, D., Veit, B. & Hake, S. Expression of maize *KNOTTED1* related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot. *Development* **120**, 405–413 (1994).
- von Arnim, A. G., Deng, X.-W. & Stacey, M. G. Cloning vectors for the expression of green fluorescent protein fusion proteins in transgenic plants. *Gene* **221**, 35–43 (1998).
- Clough, S. J. & Bent, A. F. Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
- Wysocka-Diller, J., Helariutta, Y., Fukaki, H., Malamy, J. & Benfey, P. Molecular analysis of *SCARECROW* function reveals a radial patterning mechanism common to root and shoot. *Development* **127**, 595–603 (2000).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank K. Roberts for the JIM13 antibody; M. Hahn for the CCRC-M2 antibody; B. Scheres for the QC46 marker line; J. Schiefelbein for the GL2::GUS line; A. von Arnim for the GFP plasmid; M. Aida for the pBIH vector; and M. Starz for the assistance with confocal microscopy. Multi-photon confocal images were taken with assistance by J. Feijo and N. Moreno. K.N. was supported by a fellowship from Japan Society for the Promotion of Science. This work was supported by a grant to P.N.B. from the NIH.

Correspondence and requests for materials should be addressed to P.N.B. (e-mail: philip.benfey@nyu.edu).

Archipelago regulates Cyclin E levels in *Drosophila* and is mutated in human cancer cell lines

Kenneth H. Moberg, Daphne W. Bell, Duke C. R. Wahrer, Daniel A. Haber & Iswar K. Hariharan

Massachusetts General Hospital Cancer Center, Building 149, 13th Street, Charlestown, Massachusetts 02129, USA

During *Drosophila* development and mammalian embryogenesis, exit from the cell cycle is contingent on tightly controlled down-regulation of the activity of Cyclin E–Cdk2 complexes that normally promote the transition from G1 to S phase^{1,2}. Although protein degradation has a crucial role in downregulating levels of Cyclin E, many of the proteins that function in degradation of Cyclin E have not been identified. In a screen for *Drosophila* mutants that display increased cell proliferation, we identified *archipelago*, a gene encoding a protein with an F-box and seven tandem WD (tryptophan-aspartic acid) repeats. Here we show that *archipelago* mutant cells have persistently elevated levels of Cyclin E protein without increased levels of *cyclin E* RNA. They are under-represented in G1 fractions and continue to proliferate when their wild-type neighbours become quiescent. The *Archipelago* protein binds directly to Cyclin E and probably targets it

for ubiquitin-mediated degradation. A highly conserved human homologue is present and is mutated in four cancer cell lines including three of ten derived from ovarian carcinomas. These findings implicate *archipelago* in developmentally regulated degradation of Cyclin E and potentially in the pathogenesis of human cancers.

To identify genes that restrict cell numbers and tissue growth during development, we conducted a genetic screen to identify recessive mutations that gave homozygous mutant cells even a subtle proliferative advantage over their wild-type neighbours³. Clones of homozygous mutant tissue (marked white) were generated in the eyes of heterozygous flies and their size was compared with the wild-type twin spots generated from the same recombination events. We retained flies in which there was more mutant than wild-type tissue. Of more than 23 loci identified in the screen, we obtained multiple alleles of homologues of several known human tumour-suppressor genes, including *TSC1*, *TSC2* and *PTEN*. A previously unknown locus, represented by a lethal complementation group consisting of three alleles, was named *archipelago* (*ago*).

Compared with an unmutagenized control, adult eyes mosaic for mutations in *ago* were composed mostly of mutant tissue (Fig. 1a, b). Within *ago* mutant clones, most ommatidial clusters lacked the wild-type complement of photoreceptor cells, and the distance between adjacent photoreceptor clusters was increased (Fig. 1c). Staining of apical cell profiles during pupal eye development showed that the enlarged interommatidial spaces in *ago* mutant clones contain excess cells compared with the wild type (12 ± 0 (mean $\pm$ s.d., $n = 3$) cells surrounding a single photoreceptor cluster, versus 18.5 ± 1.7 ($n = 4$) cells per cluster within *ago* clones; Fig. 1d, f). TdT-mediated dUTP nick end labelling (TUNEL) revealed no significant decrease in the extent of cell death in *ago* mutant clones (data not shown). Moreover, co-expression of the baculovirus p35 protein, which blocks caspase-dependent cell death⁴, resulted in a further increase in the number of interommatidial cells (19.5 ± 1.3 ($n = 4$) cells per cluster in the p35 background, and 39.7 ± 1.2 ($n = 3$) cells per cluster within

p35-expressing, *ago* mutant clones; Fig. 1e–g). These data suggest that loss of *ago* leads to increased cell proliferation that is partially offset by apoptosis.

Using meiotic and deletion mapping, *ago* was localized to position 64B on the left arm of chromosome 3. Sequencing candidate transcription units demonstrated that all three *ago* alleles (*ago*¹, *ago*³, *ago*⁴) have mutations in an open reading frame designated CG15010 (ref. 5). *ago* encodes a 1,326-amino-acid protein that contains an F-box and seven WD repeats in its carboxy-terminal portion (Fig. 2a). F-boxes and WD motifs are found in proteins that function as the substrate-recognition component of SCF-type ubiquitin ligase complexes^{6,7}. Within the C-terminal region, the Ago protein is most similar to that encoded by an amino-terminally truncated human expressed sequence tag (EST) previously designated FLJ11071, which we have now renamed human Ago (Fig. 2a, b). In the accompanying paper, this protein is referred to as hCdc4 (ref. 8). A more distant relative is the *Caenorhabditis elegans* gene *sel-10* (ref. 9). Two of the mutations (Ala 1,117 to Val, Gly 1,131 to Glu) are missense mutations that map to the fourth WD repeat. Each alters a residue that is conserved among the three proteins shown in Fig. 2b. The third mutation (Gln 1,195 to stop) would result in premature termination of translation early in the sixth WD repeat. As the WD repeats of F-box proteins are thought to be required for recruitment of substrates to the SCF complex⁷, it is possible that the alleles recovered in our screen may impair the interaction of Ago with specific substrates.

Because *ago* mutant cells proliferate more than wild-type cells, it seemed likely that *ago* mutations would result in increased levels of a positive regulator of the cell cycle. We generated clones of homozygous mutant tissue in eye imaginal discs and examined them for changes in cyclin levels (Fig. 3a–f). In wild-type third-instar discs (Fig. 3a), Cyclin E is expressed at varying levels and is unpatterned in cells anterior to the morphogenetic furrow, which is thought to correlate with expression at specific stages of the cell cycle in cells that are proliferating asynchronously¹⁰. A strong stripe of expression can be found immediately posterior to the furrow,

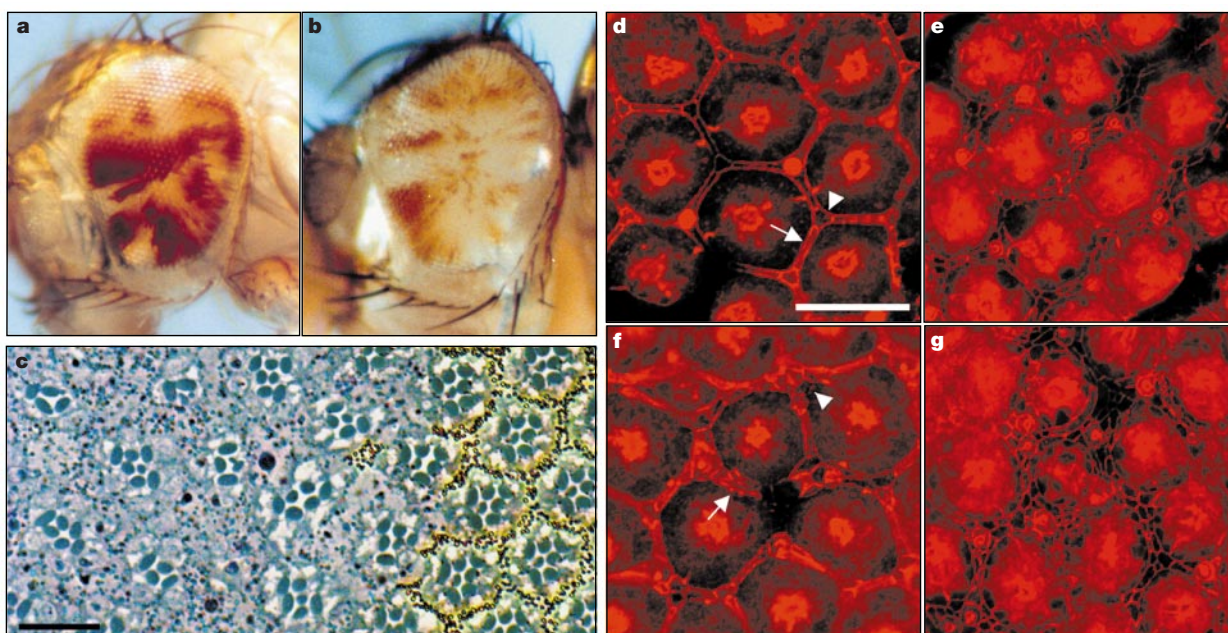


Figure 1 *ago* mutant cells have a proliferative advantage over wild-type cells. **a, b**, Adult eyes containing many homozygous clones of the parent chromosome bearing the *FRT80B* P element (**a**) and the *ago*³ allele (**b**). The mutant tissue is white because it lacks expression of the P[mini-w] element that marks the wild-type chromosome. The wild-type 'twin spot' tissue appears red. **c**, Section of an adult eye that has *ago*³ mutant clones. The mutant portions of the retina (on the left side of the image) lack the refractile pigment

granules in the interommatidial cells. Scale bar, 10 μ m. **d–g**, Apical profiles of cells in the 48-h pupal retina stained with anti-Armadillo (red). Secondary pigment cells (arrows) and tertiary pigment cells (arrowhead) are indicated. **d**, Wild type (the *FRT80B* parent chromosome); **e**, wild type expressing *GMR-p35*; **f**, clone of *ago*³ mutant cells; **g**, *ago*³ clone expressing *GMR-p35*. Scale bar, 10 μ m.

which corresponds to cells of the second mitotic wave. Cells posterior to the second mitotic wave express low levels of Cyclin E. In clones of *ago* mutant tissue anterior to the furrow, almost all cells express high levels of Cyclin E (Fig. 3b–f). Clones posterior to the furrow display mild elevations in Cyclin E levels. In contrast to the increase in Cyclin E protein, no alteration in the expression pattern of *cyclin E* RNA is observed in discs that contain many large *ago* mutant clones (Fig. 3g, h). In wild-type discs, *ago* mRNA is also expressed throughout the disc, but is expressed at particularly high levels in the morphogenetic furrow (Fig. 3i). In contrast to the results obtained with Cyclin E, the levels of Cyclin B (see Supplementary Information), Cyclin A and the SCF substrates Cubitus interruptus, Armadillo and Tramtrack are not appreciably elevated in *ago* mutant clones in the larval imaginal disc, nor are the levels of

the putative substrates Dacapo and the intracellular domain of Notch (data not shown).

Because Cyclin E promotes S-phase entry, an increase in the level of Cyclin E can perturb regulation of the cell cycle. To examine the proliferative properties of *ago* mutant cells, we generated *ago* clones in the wing disc of third-instar larvae and compared their DNA content with that of wild-type cells from the same imaginal discs (Fig. 4e). In mutant clones, a smaller fraction of cells (21.4%) is found in G1 when compared with wild-type cells (36.8%). The proportion of cells found in S phase and in G2/M is increased. These alterations are extremely similar to those elicited by the overexpression of Cyclin E¹¹.

We also examined the effect of *ago* mutations on the patterns of cell proliferation *in vivo*. Cells anterior to the morphogenetic furrow

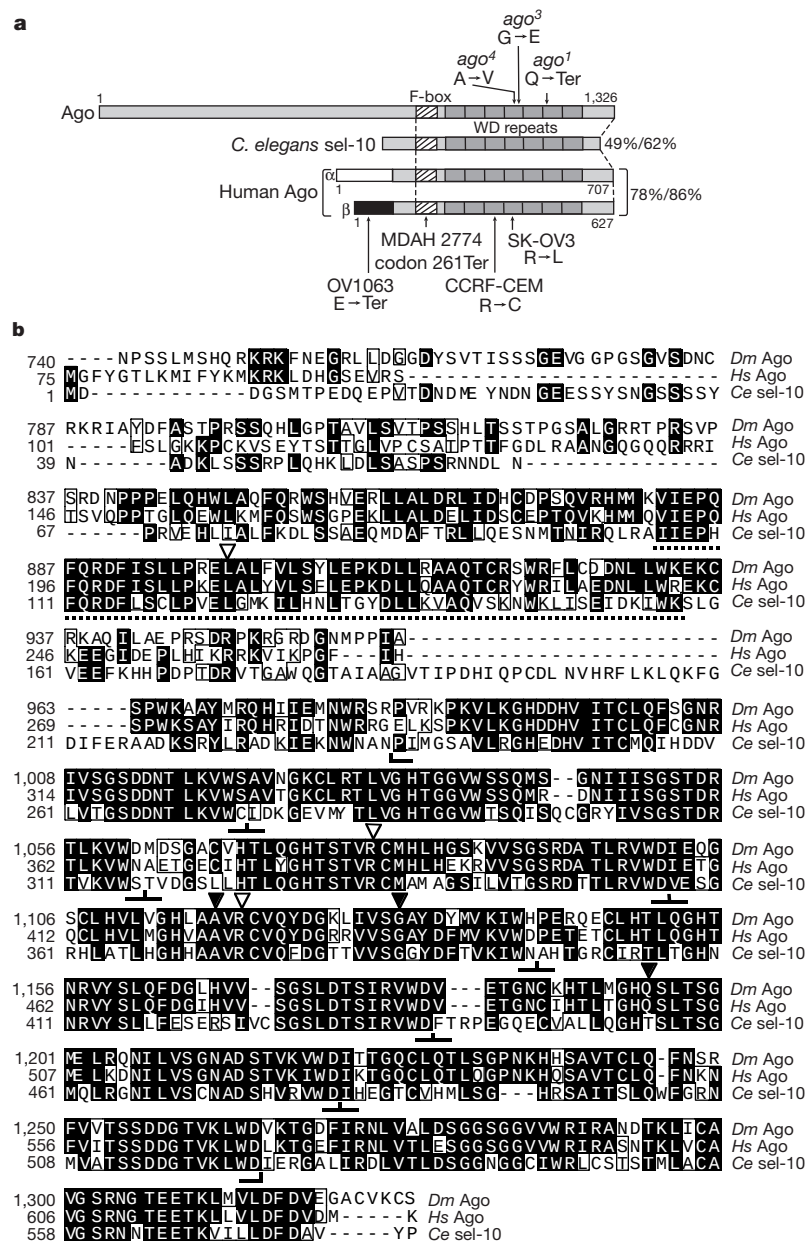


Figure 2 *ago* encodes a protein with an F-box and WD repeats, and its human orthologue is mutated in cancer cell lines. **a**, The domain structure of the Ago protein showing the F-box (hatched), WD-repeat region (dark grey) and the mutations identified in the *ago* alleles. The two N-terminal forms of human Ago are indicated (α , white; β , black) and the positions of mutations in cancer cell lines are indicated. The degree of conservation in the F-box/WD region between *Drosophila* and human Ago and *C. elegans sel-10* is indicated

(identity/similarity). **b**, Amino-acid comparison between *Drosophila* (*Dm*) Ago, the β form of human (*Hs*) Ago, and *C. elegans* (*Ce*) sel-10. The F-box (dotted underline), seven WD repeats (brackets), and location of mutations in *Drosophila* (solid arrowhead) and human cell lines (open arrowhead) are indicated (also see Supplementary Information). Identical (black box) and similar (open box) residues are indicated.

in the larval eye disc proliferate asynchronously. It is therefore difficult to visualize differences in rates of cell proliferation in mutant clones at this stage of eye development. In contrast, very few cells proliferate in the wild-type pupal retina. The bristle precursor cell is the only mitotically active cell type detected during this stage; it divides twice during the mid-pupal phase to generate the four cells of the 'bristle complex'¹². Levels of Cyclin E rapidly decrease after these divisions. In *ago* mutant clones, elevated

levels of Cyclin E are detected in the four cells of the bristle complex well after the levels in the corresponding cells of adjacent wild-type ommatidia have declined (Fig. 4a–c). Some of these *ago* mutant cells also continue to incorporate 5-bromodeoxyuridine (BrdU), suggesting that they continue to cycle after the corresponding cells in adjacent wild-type tissue have exited from the cell cycle (Fig. 4d). Such additional divisions are likely to contribute to the increased number of interommatidial cells observed in the pupal retina. Thus,

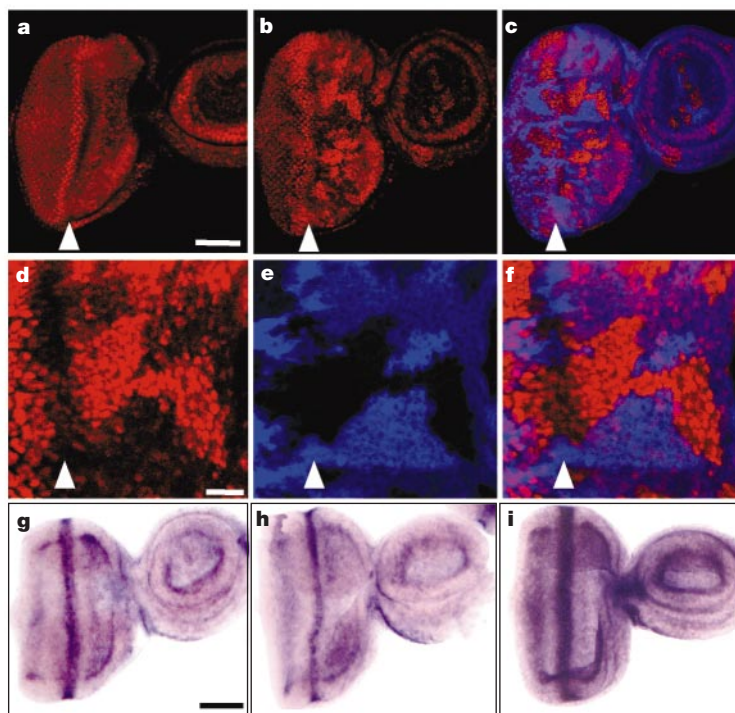


Figure 3 *ago* inhibits accumulation of Cyclin E by a post-transcriptional mechanism. Third-instar eye imaginal discs, anterior to the right. **a–f**, Anti-Cyclin E staining (red) in wild-type (**a**) or mosaic *ago* (**b–f**) larval eye discs; *ago* mutant tissue is marked by the absence of anti- β -galactosidase (β -gal) staining (blue) in **c**, **e** and **f**. Arrowheads denote

the morphogenetic furrow. RNA *in situ* analysis shows that levels of *cyclin E* RNA are the same in wild-type (**g**) discs and discs containing large *ago*³ clones (**h**). Expression of *ago* RNA in a wild-type disc (**i**). Scale bars: **a**, **g**, 50 μ m; **d**, 10 μ m.

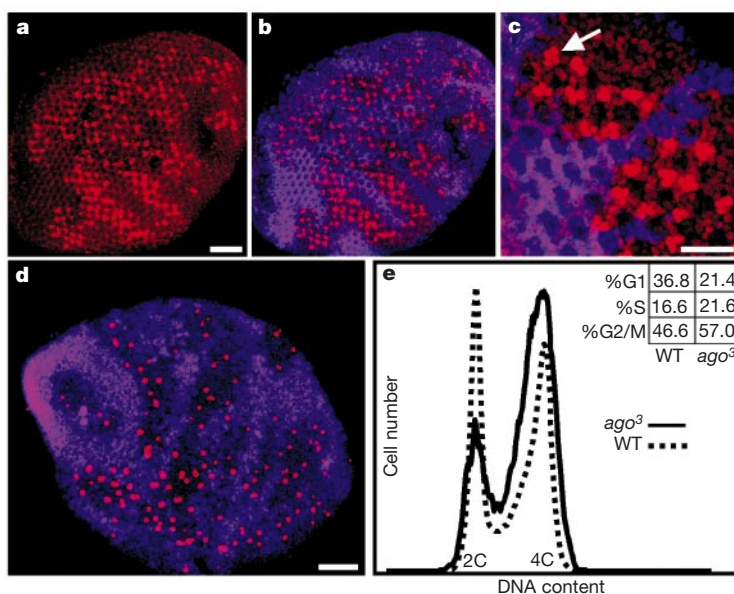


Figure 4 Loss of *ago* alters regulation of the cell cycle. **a–d**, 24-h pupal discs containing *ago* mutant clones; *ago* mutant tissue is marked by the absence of anti- β -gal staining (blue) in **b–d**. **a–c**, Anti-Cyclin E staining in red. Arrow in **c** indicates a cluster of four bristle-complex nuclei within an *ago* mutant clone, which stain brightly for Cyclin E.

d, Anti-BrdU shows S-phase nuclei in 24-h pupal eye discs. **e**, Analysis of cell cycle phasing by flow cytometry in wild-type (WT) and *ago*³ mutant cells generated in larval wing discs. Scale bars: **a**, **d**, 50 μ m; **c**, 20 μ m.

the persistence of Cyclin E in *ago* mutant cells disrupts exit from the cell cycle in a manner similar to that elicited by Cyclin E overexpression¹³.

The simplest explanation of the role of *ago* in cell cycle control is that Ago binds to Cyclin E and targets it for ubiquitin-mediated degradation. We therefore looked for genetic and physical interactions between Ago and Cyclin E. A genetic interaction was observed between *ago* and the *cyclin E^{EP}* allele, which reduces the levels of cyclin E transcription in the developing eye¹⁴. The rough-eye phenotype of *cyclin E^{EP}* flies is suppressed in flies that are also heterozygous for a mutation in *ago* (Fig. 5a–c). In addition, *ago* mutations dominantly suppress the small-eye phenotype produced by *eyGAL4*-driven overexpression of the cyclin-dependent kinase inhibitor *dacapo* (data not shown), which has been shown to reduce Cyclin E–Cdk2 activity. Thus flies that are heterozygous for mutant

alleles of *ago* are likely to have increased levels of Cyclin E.

To test for a direct physical interaction between Archipelago and Cyclin E, we expressed the portion of Archipelago containing the F-box and WD repeats as a protein fused to glutathione S-transferase (GST: GST–AgoΔN) and evaluated its ability to bind Cyclin E protein in lysates of S2 cells transfected with *cyclin E* and *cdk2*. We also generated versions of GST–AgoΔN that contained the mutations found in the *ago¹* and *ago³* alleles. Binding was readily detected with the wild-type version of GST–AgoΔN and was greatly reduced with both mutant versions (Fig. 5d). Thus the ability of Archipelago to bind Cyclin E *in vitro* correlates with its ability to downregulate Cyclin E levels *in vivo*.

Elevated levels of Cyclin E are observed in a subset of human tumours, notably those of the breast and ovary^{15,16}. We therefore examined human cancer cell lines and primary tumours for mutations in the human *ago* orthologue. The EST FLJ11071 encoding human *ago* is incomplete at the 5' end. Using exon-prediction programmes, we were able to identify additional sequences that indicate the presence of least two alternative 5' exons, and confirmed the structure of the putative transcription unit using polymerase chain reaction with reverse transcription (RT-PCR). When these sequences are included, the human *ago* gene encodes proteins of 707 and 627 amino acids, which we have termed α and β, respectively (Fig. 2a). The α and β forms differ only in their N termini, and so are identical in the putative functional domain containing the F-box and WD repeats. The entire coding region was sequenced by RT-PCR from 17 breast cancer cell lines, 8 primary breast tumours and 35 other cell lines derived from tumours of the ovary, lung, colon, kidney, brain and bone as well as melanomas and leukaemias.

Mutations were detected in four of the cell lines (see Supplementary Information). Two cell lines were heterozygous for missense mutations in conserved residues in the WD repeats (Fig. 2a). The cell line derived from T-cell acute lymphoblastic leukaemia, CCRF-CEM, contains an arginine-to-cysteine change within the third WD repeat at codon 385, and the SK-OV3 ovarian cancer cell line contains an arginine-to-leucine change within the fourth WD repeat at codon 425. Significantly, the latter mutation alters an amino acid only two residues away from that mutated in the *Drosophila ago¹* allele (Fig. 2b). The mutation in the ovarian cell line MDAH 2774 disrupts a splice acceptor site, resulting in a protein that lacks part of the F-box and all of the WD repeats. It is worth noting that, as in *Drosophila* (Fig. 5a–c), these heterozygous mutations in human *ago* may give cells a proliferative advantage. Most significantly, the ovarian cell line OV1063 is homozygous for a truncating mutation close to the N terminus of the β form that most probably inactivates this isoform. Thus three of ten ovarian cancer cell lines examined have mutations in human *ago*, indicating that this gene might represent an important mutational target in ovarian cancer. Interestingly, almost 20% of ovarian cancers have been shown to express increased levels of *cyclin E* mRNA, often the result of gene amplification¹⁶. Our findings indicate that mutations in human *ago* could represent an alternative mechanism for deregulating Cyclin E in these tumours.

Given that Cyclin E can be deregulated by a variety of mechanisms in ovarian tumours, we examined Cyclin E RNA and protein expression in a panel of seven ovarian cancer cell lines. Two of seven cell lines (OVCAR3 and OVCAR4) were found to express comparatively high levels of Cyclin E mRNA (Fig. 5f). Among the remaining cell lines, which expressed low levels of Cyclin E mRNA, two lines showed especially high levels of Cyclin E protein (Fig. 5e). One of these is the cell line with the homozygous mutation (OV1063) in human Agoβ. The other cell line (OVCAR8) did not have a mutation in the coding region of human Ago, indicating a possible mutation in another regulator of levels of Cyclin E protein.

We have shown that the *Drosophila* F-box protein Archipelago regulates the levels of Cyclin E protein post-transcriptionally during

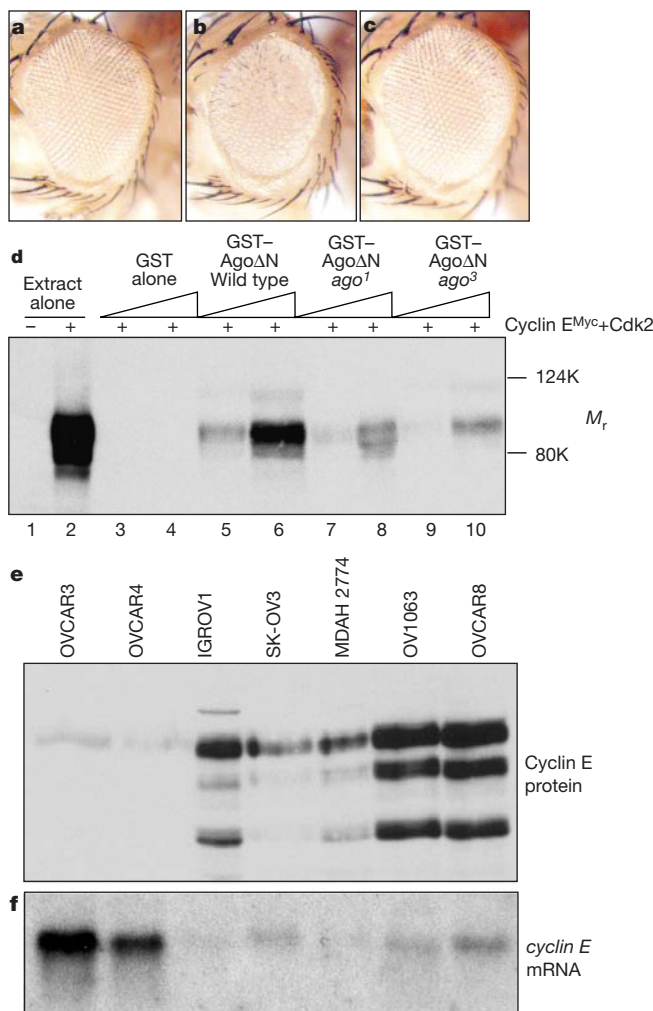


Figure 5 Genetic and physical interaction between Archipelago and Cyclin E. **a–c**, Images of adult eyes of the *w; FRT80B* (wild type) (**a**), *w; cyclin E^{EP}; FRT80B/+* (**b**), and *w; cyclin E^{EP}; FRT80B, ago³/+* (**c**) genotypes. **d**, *In vitro* binding of a GST–Ago fusion protein to Cyclin E. Extracts derived from S2 cells expressing *Drosophila* Cyclin E/Cdk2 were assayed for binding to the portion of wild-type (lanes 5 and 6) or mutant Archipelago proteins (lanes 7–10) containing the F-box and WD repeats, expressed as GST fusions and immobilized on glutathione beads. The mutant forms correspond to those found in the *ago¹* and *ago³* alleles. Lane 2 represents 5% of the lysate used in each binding reaction. Cyclin E bound to GST–Archipelago beads was detected by western blotting using an antibody that detects the Myc tag on the N terminus of Cyclin E. Units on right are relative molecular mass (*M_r*), in thousands. **e**, Cyclin E protein derived from thymidine-blocked ovarian cancer cell lines detected by western blotting using anti-Cyclin E antibody. **f**, *cyclin E* RNA from cancer cell lines detected by northern blotting.

normal development. The mutations that disrupt its function *in vivo* reduce binding to Cyclin E and delay a decay in levels of Cyclin E. It has previously been shown that Cyclin E can be degraded in mammalian cells by direct interaction with a cullin^{17–19}. Our findings, together with the observation that mutations in the *C. elegans* genes *cull1* and *lin-23* (which encode a cullin and an F-box protein respectively) have increased cell divisions^{20,21}, highlight the importance of SCF-mediated degradation in regulating cell proliferation through Cyclin E. Because *ago* RNA is expressed in a dynamic pattern, our results indicate that degradation of Cyclin E is not constitutive *in vivo*. Dynamic expression of Ago provides another mechanism by which cyclin/cdk activity and cell proliferation can be regulated during development. Finally, we implicate impaired proteolysis of Cyclin E in the pathogenesis of human cancers. □

Methods

Fly stocks

All crosses were conducted at 25 °C. *w; FRT80B* males were mutagenized with ethylmethanesulphonate (EMS), then crossed to *y w eyFLP;FRT80B P[mini-w, arm-LacZ]* virgin females (stocks a gift of J. Treisman). Males with more white than red eye tissue were selected and maintained as balanced stocks. Alleles of *archipelago* isolated were *ago¹*, *ago³* and *ago⁴*. Other stocks were *y w hsFLP;FRT80B P[πMyc] P[w y]*, *w;FRT80B P[mini-w] P[UbiGFP]/TM6B* (a gift of B. Edgar), *GMR-p35* (a gift of K. White) and *w; cycE^{EP}* (a gift of H. Richardson).

Microscopy, immunohistochemistry, flow cytometry

Adult eyes were photographed submerged in mineral oil. Imaginal disc tissue of the indicated genotypes was fixed and stained for Cyclin E and β-galactosidase as described previously⁷. Images were collected on a Carl Zeiss Axiovert 100M Confocal microscope. The mouse monoclonal antibody to *Drosophila* type I Cyclin E and the Cyclin B antibody were gifts of H. Richardson and C. Lehner, respectively. The antisense *Drosophila cyclin E* probe was derived from full-length type I *cyclin E* complementary DNA. The antisense *ago* probe was derived from a full-length cDNA. Flow cytometry on third-instar larval wing discs was performed as described previously³. For *ago* loss-of-function FACS analysis, the following genotype was used: *y w hsFLP; FRT80B ago³/FRT80B P[mini-w] P[UbiGFP]*.

Molecular biology

For GST-fusion proteins, PCR fragments corresponding to the C-terminal 660 amino acids of the wild-type or *ago* mutant open reading frames were cloned in-frame into the pGEX-2T vector (Amersham Pharmacia). Following induction, equal amounts of intact GST–AgoΔN fusion proteins were incubated with 100 μg of S2 whole-cell extract from cells transfected using the CELLECTIN reagent (Gibco BRL) with *Drosophila cdk2* and Myc-epitope-tagged *Drosophila* type I *cyclin E* cDNAs cloned into pIE1–4 insect expression vectors (Novagen). Myc-tagged cyclin E protein was detected in western blots using the 9E10 anti-Myc tag monoclonal antibody; human Cyclin E was detected in lysates of human ovarian cancer cell lines synchronized in G1/S by incubation for 36 h in 2 mM thymidine with the HE12 anti-cyclin E monoclonal antibody (both antibodies were a gift of E. Harlow). For northern analysis, 10 μg of total cellular RNA was probed with a 0.9-kilobase *Bam*H1–*Xmn*I fragment of the human *cyclin E* cDNA.

Characterization of human *ago*

The α and β forms of human *ago* were identified in BAC067826 by exon-prediction programmes. The existence of both forms was confirmed by RT-PCR. An in-frame termination codon is present 75 nucleotides upstream from the initiating ATG in the β cDNA. We were unable to locate an upstream in-frame termination codon in the α cDNA. These sequences have been deposited in GenBank. Human *ago* was amplified by RT-PCR in six overlapping fragments. PCR products were resolved by gel electrophoresis and sequenced directly with the BigDye Terminator kit (Applied Biosystems) and analysed on an ABI300 genetic analyser. In addition to eight primary tumours, the cancer cell lines analysed were: breast (MCF7ADR, MDAMB435, T47D, BT483, MDAMB436, MDAMB453, MDAMB468, MDAMB415, MDAMB231, MDAMB175, MDAMB157, HS157, HS467T, HS496T, HS578T, UACC893, BT549), ovarian (ES-2, IGROV-1, MDAH2774, OV1063, OVCAR3, OVCAR4, OVCAR5, OVCAR8, SKOV3, SW626), lung (NCIH460, NCI522, HOP92), central nervous system (SF295, SNB19, U251), leukaemia (CCRF-CEM, K562, MOLT4, RPMI-8226, SR), colon (COLO205, HCT116, HCT15), renal (786-0, ACHN, CAKI-1, SN12C, UO31), melanoma (LOXMVII, M14, SKMEL2, UACC62) and osteosarcoma (U2OS, SAOS2). The wild-type controls were EBV-immortalized cell lines from normal individuals²².

Received 29 May; accepted 27 August 2001.

- Knoblich, J. A. *et al.* Cyclin E controls S phase progression and its down-regulation during *Drosophila* embryogenesis is required for the arrest of cell proliferation. *Cell* **77**, 107–120 (1994).
- Kiyokawa, H. & Koff, A. Roles of cyclin-dependent kinase inhibitors: lessons from knockout mice. *Curr. Top. Microbiol. Immunol.* **227**, 105–120 (1998).
- Tapon, N., Ito, N., Dickson, B. J., Treisman, J. E. & Hariharan, I. K. The *Drosophila* tuberous sclerosis complex gene homologs restrict cell growth and cell proliferation. *Cell* **105**, 345–355 (2001).
- Hay, B. A., Wolff, T. & Rubin, G. M. Expression of baculovirus P35 prevents cell death in *Drosophila*. *Development* **120**, 2121–2129 (1994).

- Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
- Bai, C. *et al.* SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. *Cell* **86**, 263–274 (1996).
- Skowyr, D., Craig, K. L., Tyers, M., Elledge, S. J. & Harper, J. W. F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell* **91**, 209–219 (1997).
- Strohmaier, H. *et al.* Human F-box protein hCdc4 targets cyclin E for proteolysis and is mutated in a breast cancer cell line. *Nature* **413**, 316–322 (1998).
- Hubbard, E. J., Wu, G., Kitajewski, J. & Greenwald, I. sel-10, a negative regulator of lin-12 activity in *Caenorhabditis elegans*, encodes a member of the CDC4 family of proteins. *Genes Dev.* **11**, 3182–3193 (1997).
- Horsfield, J., Penton, A., Secombe, J., Hoffman, F. M. & Richardson, H. decapentaplegic is required for arrest in G1 phase during *Drosophila* eye development. *Development* **125**, 5069–5078 (1998).
- Neufeld, T. P., de la Cruz, A. F., Johnston, L. A. & Edgar, B. A. Coordination of growth and cell division in the *Drosophila* wing. *Cell* **93**, 1183–1193 (1998).
- Perry, M. M. Further studies on the development of the eye of *Drosophila melanogaster*. II: The interommatidial bristles. *J. Morphol.* **124**, 249–262 (1968).
- Richardson, H., O'Keefe, L. V., Marty, T. & Saint, R. Ectopic cyclin E expression induces premature entry into S phase and disrupts pattern formation in the *Drosophila* eye imaginal disc. *Development* **121**, 3371–3379 (1995).
- Secombe, J., Pispis, J., Saint, R. & Richardson, H. Analysis of a *Drosophila* cyclin E hypomorphic mutation suggests a novel role for cyclin E in cell proliferation control during eye imaginal disc development. *Genetics* **149**, 1867–1882 (1998).
- Keyomarsi, K. & Herliczek, T. W. The role of cyclin E in cell proliferation, development and cancer. *Prog. Cell Cycle Res.* **3**, 171–191 (1997).
- Courjal, F. *et al.* Cyclin gene amplification and overexpression in breast and ovarian cancers: evidence for the selection of cyclin D1 in breast and cyclin E in ovarian tumors. *Int. J. Cancer* **69**, 247–253 (1996).
- Singer, J. D., Gurian-West, M., Clurman, B. & Roberts, J. M. Cullin-3 targets cyclin E for ubiquitination and controls S phase in mammalian cells. *Genes Dev.* **13**, 2375–2387 (1999).
- Dealy, M. J. *et al.* Loss of Cull1 results in early embryonic lethality and dysregulation of cyclin E. *Nature Genet.* **23**, 245–248 (1999).
- Wang, Y. *et al.* Deletion of the Cull1 gene in mice causes arrest in early embryogenesis and accumulation of cyclin E. *Curr. Biol.* **9**, 1191–1194 (1999).
- Kipreos, E. T., Lander, L. E., Wing, J. P., He, W. W. & Hedgecock, E. M. cul-1 is required for cell cycle exit in *C. elegans* and identifies a novel gene family. *Cell* **85**, 829–839 (1996).
- Kipreos, E. T., Gohel, S. P. & Hedgecock, E. M. The *C. elegans* F-box/WD-repeat protein LIN-23 functions to limit cell division during development. *Development* **127**, 5071–5082 (2000).
- FitzGerald, M. G. *et al.* Heterozygous ATM mutations do not contribute to early onset of breast cancer. *Nature Genet.* **15**, 307–310 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We are indebted to C. Paulding for help with exon predictions, to S. Schelble for technical assistance, and to C. Pfeiffer, N. Dyson and O. Iliopoulos for comments on the manuscript. I.K.H. was funded by grants from the National Institutes of Health and the American Cancer Society (ACS) and K.H.M. was funded by a postdoctoral fellowship from the ACS.

Correspondence and requests for materials should be addressed to I.K.H. (e-mail: hariharan@helix.mgh.harvard.edu). The sequences for human Ago have been deposited in GenBank under accession numbers AF411971 (α form) and AF411972 (β form).

Human F-box protein hCdc4 targets cyclin E for proteolysis and is mutated in a breast cancer cell line

Heimo Strohmaier*, Charles H. Spruck*, Peter Kaiser*, Kwang-Ai Won*, Olle Sangfelt† & Steven I. Reed*

* Department of Molecular Biology, MB-7, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA

† Department of Oncology/Pathology, CCK R8:03, Karolinska Hospital, 171 76 Stockholm, Sweden

Cyclin E, one of the activators of the cyclin-dependent kinase Cdk2, is expressed near the G₁–S phase transition and is thought to be critical for the initiation of DNA replication and other S-phase functions^{1–3}. Accumulation of cyclin E at the G₁–S boundary is achieved by periodic transcription coupled with regulated proteolysis linked to autophosphorylation of cyclin E⁴. The proper timing and amplitude of cyclin E expression seem to

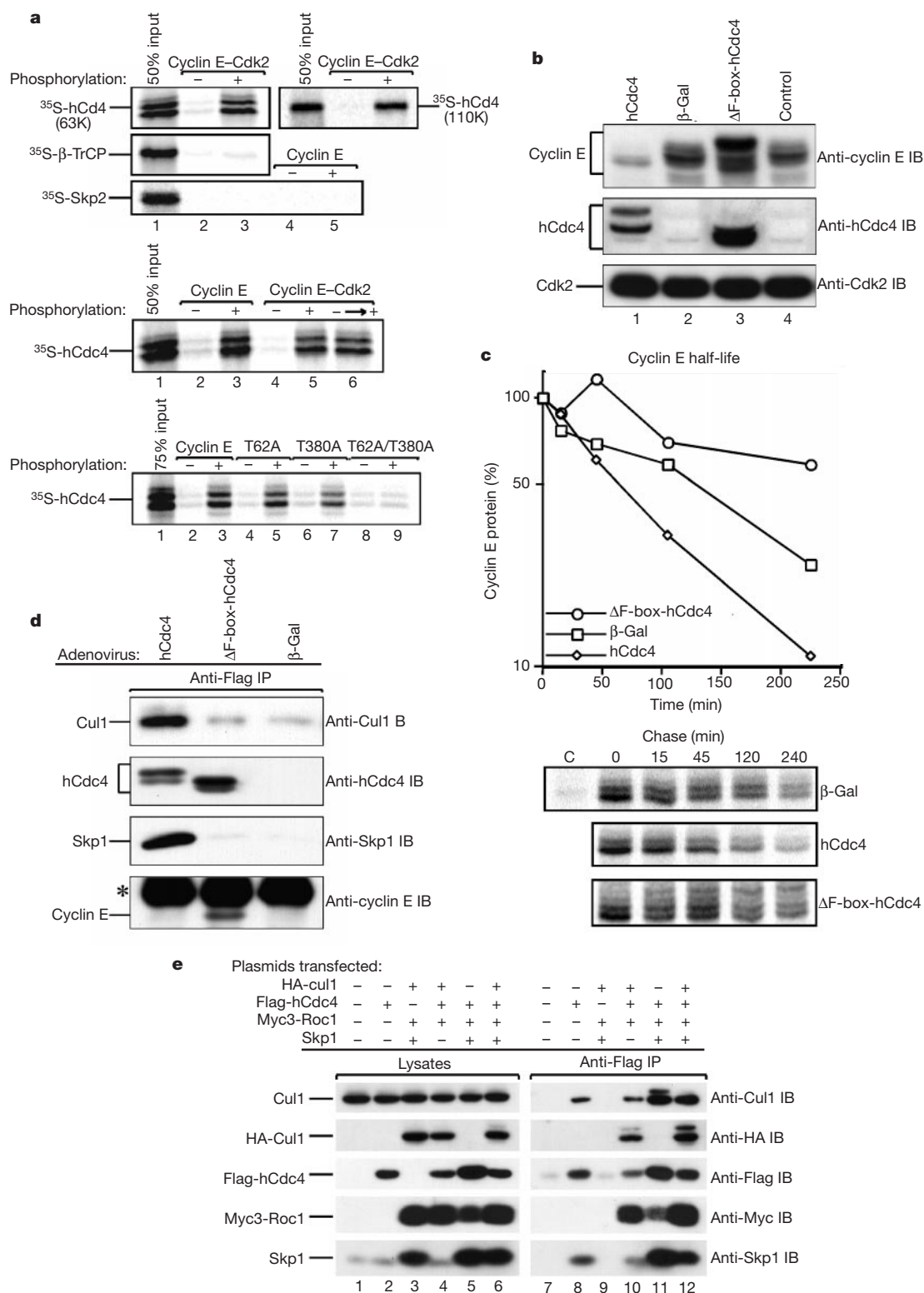


Figure 2 Human Cdc4 assembles into SCF complexes with Cul1, Skp1 and Roc1 *in vivo*, and regulates cyclin E turnover through specific association with phosphorylated cyclin E. **a**, *In vitro* translated, ³⁵S-labelled hCdc4 (63K and 110K forms), β-TrCP and Skp2 were assayed for binding to either free or Cdk2-bound GST-tagged cyclin E purified from SF9 insect cells on glutathione beads (lanes 1–5, top and middle panels). Dephosphorylation of cyclin E was performed after purification with λ phosphatase. ³⁵S-labelled hCdc4 was also tested for binding to cyclin E that had been dephosphorylated followed by rephosphorylation by its associated kinase, Cdk2, in the presence of 1 mM ATP (lane 6, middle panel), and for binding to single phosphorylation-site mutants (T62A, T380A) as well as the double mutant (T62A/T380A) (bottom panel). **b**, Thymidine-

arrested KB cells were transduced with recombinant adenoviruses expressing hCdc4 (lane 1), a ΔF-box hCdc4 (lane 3) or a control virus (β-galactosidase) (lane 2). Lane 4 is an uninfected control. **c**, ³⁵S-methionine pulse-chase analysis of cyclin E in adenovirally transduced KB cells described in **b**. **d**, Immunoprecipitation (IP) of adenovirally transduced wild-type and mutant hCdc4 from KB cells described in **b**, and analysis of coprecipitated proteins by western blotting. In the bottom panel, the asterisk corresponds to the immunoglobulin-γ (IgG) heavy chain. **e**, 293T cells were transfected with the indicated plasmids and lysates were used for anti-Flag immunoprecipitations. Immune complexes (lanes 7–12) or crude lysates (lanes 1–6) from each transfection were analysed for the presence of Cul1, Skp1, Roc1 and hCdc4 by immunoblotting (IB).

was unchanged in a *cdc4/sic1* double mutant relative to the *cdc4* mutant (Fig. 1a), confirming that Cdc4 is indeed the critical F-box protein for cyclin E degradation in yeast.

Close scrutiny of half-life data obtained for the T380A mutant and comparison with data for wild-type cyclin E in SCF mutants suggested that the T380A mutant may still be susceptible to SCF-mediated ubiquitination and proteolysis. Accordingly, mutations were constructed at other potential phosphorylation sites. We found that the T62A mutation rendered cyclin E slightly more stable than

the wild type (Fig. 1b). Significantly, the double mutant (T62A/T380A) was more stable than the T380A mutant (Fig. 1b). This suggests that T62 is a secondary phosphorylation site involved in ubiquitination and turnover of cyclin E.

We found a human expressed-sequence tag (EST) in the EST database of GenBank (<http://www.ncbi.nlm.nih.gov>) that, when translated, had significant homology to yeast Cdc4 (Fig. 1c). Analysis of hCdc4 complementary DNAs from a number of cell lines and the genomic structure of the *hCDC4* locus indicated that

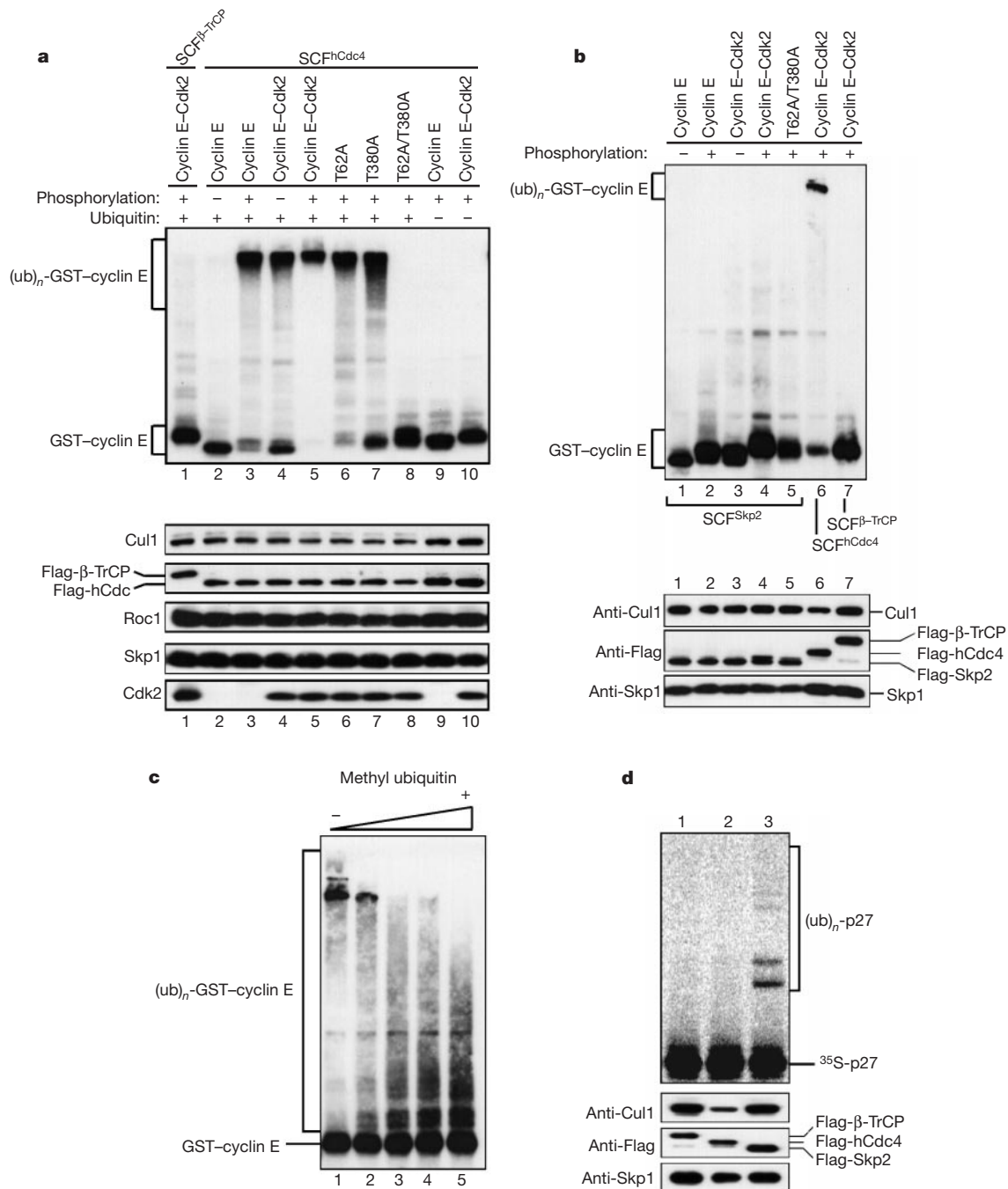


Figure 3 SCF^{hCdc4} ubiquitinates cyclin E in a phosphorylation-dependent manner *in vitro*. **a**, Anti-Flag immunoprecipitates from 293T cells transfected with either SCF^{hCdc4} (lanes 2–10) or SCF^{β-TrCP} (lane 1) were assayed for the ability to ubiquitinate cyclin E and phosphorylation-site mutants (top panel). The bottom panel shows anti-Flag immunoprecipitates. **b**, Comparison of immunoprecipitated SCF^{Skp2} (lanes 1–5), SCF^{hCdc4}

(lane 6) and SCF^{β-TrCP} (lane 7) for the ability to ubiquitinate cyclin E (top panel). ub, ubiquitin. Immunoblots for Cul1, Skp1 and Flag-tagged F-box proteins are shown (bottom panel). **c**, Effect of increasing concentrations of methylated ubiquitin on the mobility of poly-ubiquitinated cyclin E species. **d**, Comparison of the ability of immunoprecipitated SCF complexes to ubiquitinate p27^{Kip1}.

two alternatively spliced variants exist that encode proteins with different amino termini (Fig. 1c, top two lines). A search for related proteins in other species suggested that homologues exist in *Drosophila melanogaster* as well as *Caenorhabditis elegans* (Fig. 1c). *In vitro* translation of the *hCDC4* cDNA produced a polypeptide with a relative molecular mass of about 63,000 (M_r 63K). However, further analysis of *hCDC4* transcripts and genomic structure indicated that two alternative forms of the protein exist. The 63K polypeptide derived from the initial hCdc4 EST is a slightly truncated version of a 69K species, which appears to be quite tissue specific (see below). The prevalent species in most tissues, consisting of 707 amino acids, runs aberrantly on SDS gels at an apparent M_r of 110K (see below). To generate the 110K form, one large upstream coding exon is substituted for the first exon of the 69K species (data not shown). We found that expression of the 110K form of hCdc4 could partially rescue a *cdc4* mutation in the yeast *Saccharomyces cerevisiae* (data not shown), consistent with these proteins being functional and structural homologues.

To determine whether hCdc4 interacts specifically with phosphorylated cyclin E, hCdc4 translated *in vitro* (the 63K and 110K forms) was incubated with glutathione beads bound to either free glutathione S-transferase (GST)–cyclin E complexes or GST–cyclin E–Cdk2 complexes. In parallel samples, cyclin E was either phosphorylated or dephosphorylated. hCdc4 binds to phosphorylated

free or Cdk2-bound cyclin E, but not to dephosphorylated cyclin E, regardless of Cdk2 binding (Fig. 2a). In contrast, *in vitro* translated β -TrCP, another human WD40-repeat-containing F-box protein, bound to neither phosphorylated nor dephosphorylated cyclin E (Fig. 2a). It has been suggested that the human F-box protein Skp2, which contains leucine-rich repeats rather than WD40 repeats, targets cyclin E for ubiquitin-dependent degradation^{16,17}. However, *in vitro* translated Skp2 bound to neither phosphorylated nor dephosphorylated cyclin E, either free or bound to Cdk2 (Fig. 2a). hCdc4 (63K) translated *in vitro* was assayed further for the ability to bind to cyclin E phosphorylation site mutant proteins. Cyclin E (T380A) was subjected to phosphorylation and dephosphorylation and assayed for hCdc4 binding as had been done for wild-type cyclin E. Binding of this mutant protein was reduced but not eliminated (Fig. 2a). However, on the basis of our analysis presented in Fig. 1b, we assayed both the T62A (single) and T62A/T380A (double) mutants for hCdc4 binding. Cyclin E (T62A) was only slightly reduced in hCdc4 binding (Fig. 2a). However, cyclin E (T62A/T380A) was completely defective in binding (Fig. 2a), consistent with the *in vivo* half-life data (Fig. 1b).

To determine whether expression of hCdc4 *in vivo* has an impact on cyclin E turnover, KB cells were transduced with an hCdc4 recombinant adenovirus. Endogenous levels of cyclin E were

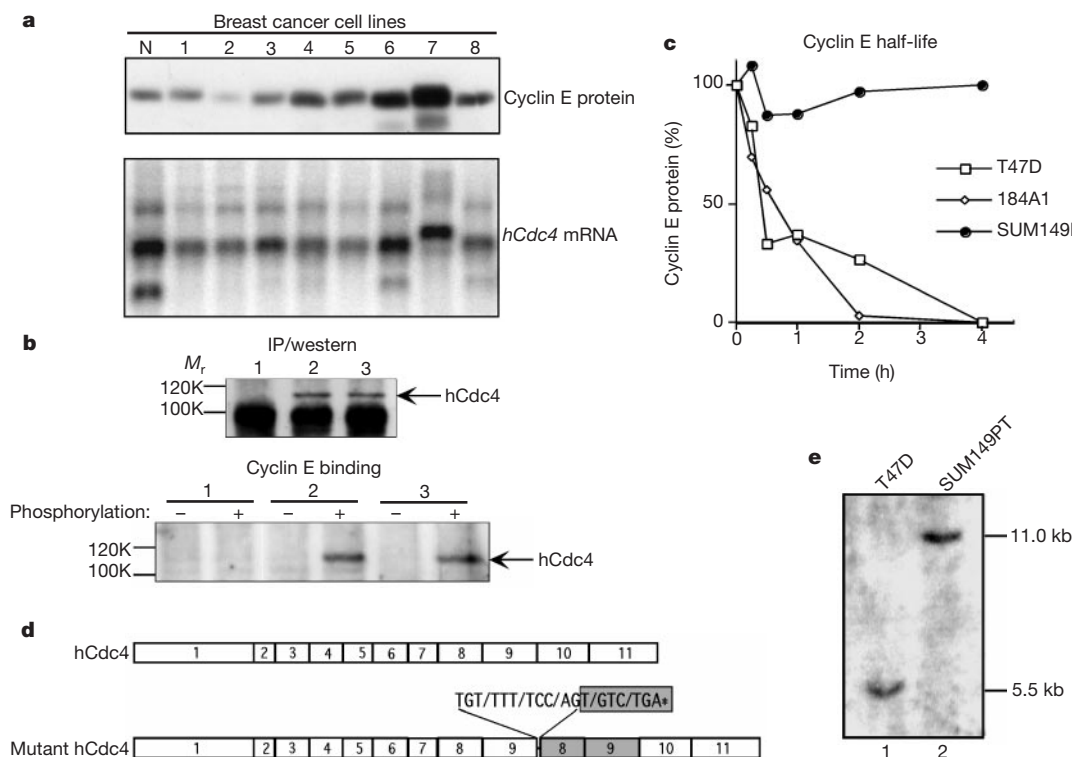


Figure 4 Aberrant hCdc4 mRNA, loss of hCdc4 protein and loss of heterozygosity in cell lines derived from breast cancer with high cyclin E expression. **a**, Eight randomly chosen breast-cancer-derived cell lines (lanes 1–8) and a breast epithelial cell line (184A1, lane N) were analysed for cyclin E expression by immunoblotting (top panel), and for hCdc4 transcripts by northern blot analysis (bottom panel). Cell lines: lane 1, MDA-MB-435S; lane 2, T47D; lane 3, BT-549; lane 4, ZR-75-1; lane 5, MDA-MB-436; lane 6, MDA-MB-157; lane 7, SUM149PT; lane 8, SK-BR-3. **b**, The breast epithelial cell line 184A1 (lane 2) and the breast-cancer-derived cell lines SUM149PT (lane 1) and T47D (lane 3) were analysed for expression of hCdc4 protein by anti-hCdc4 immunoprecipitation (IP) followed by immunoblotting using specific anti-hCdc4 antibodies (top panel) or by incubating crude lysates prepared from these cell lines with either

dephosphorylated (control) or phosphorylated GST–cyclin E immobilized on glutathione beads followed by SDS–PAGE and western blotting with anti-hCdc4 antibodies (bottom panel). The heavy band migrating ahead of hCdc4 in the immunoprecipitation–immunoblot experiment corresponds to IgG heavy chain–light chain heterodimers. **c**, ³⁵S-methionine pulse-chase analysis was performed to measure the turnover rate of cyclin E in the indicated cell lines. **d**, Structure of SUM149PT hCdc4 cDNA. Exons are numbered. Shaded exons are duplicated in tandem resulting from a tandem genomic duplication of the region containing exons 8 and 9. Spliced intronic sequences in the cDNA are shown, which lead to a chain termination at the beginning of the duplicated exon 8. **e**, Southern blot analysis of genomic DNA from breast-cancer-derived cell lines T47D and SUM149PT.

dramatically reduced compared with control adenovirus transductions (Fig. 2b). Conversely, when an adenovirus expressing an F-box-deleted (Δ F-box) and thereby, most probably, dominant negative hCdc4 allele, was transduced into the same cell line, a significant accumulation of cyclin E was observed (Fig. 2b). In addition, the bulk of accumulated cyclin E was hyperphosphorylated on the basis of SDS–polyacrylamide gel electrophoresis (SDS–PAGE) mobility (Fig. 2b), consistent with a block in the degradation specifically of phosphorylated cyclin E. 35 S-methionine pulse-chase experiments were performed on parallel adenoviral transductions (Fig. 2c). Transduction of wild-type hCdc4 led to a decrease in cyclin E half-life, whereas transduction of the dominant negative hCdc4 allele led to an increase in cyclin E half-life (Fig. 2c). Thus hCdc4 levels are rate limiting for cyclin E turnover. To confirm that the Δ F-box version of hCdc4 indeed had the appropriate characteristics to behave as a dominant negative, we showed by immunoprecipitation that it binds specifically to phosphorylated cyclin E *in vivo* but not to components of SCF (Fig. 2d).

To confirm that wild-type hCdc4 is part of an SCF complex, Flag-tagged hCdc4 was introduced into 293T cells by transfection. Analysis of anti-Flag immunoprecipitates indicated that hCdc4 is associated with both endogenous and co-transfected core components of human SCF (Fig. 2e).

To determine whether SCF^{hCdc4} can ubiquitinate cyclin E in a phosphorylation-dependent manner, expression plasmids for Flag-hCdc4, as well as the other three components of SCF, were cotransfected into 293T cells. Anti-Flag immunoprecipitates were then tested for their ability to ubiquitinate phosphorylated cyclin E. Immunoprecipitated SCF^{hCdc4} was capable of efficiently ubiquitinating phosphorylated cyclin E, either free or bound to Cdk2 (Fig. 3a, b). Addition of methylated ubiquitin increased the mobility of the cyclin E derivatives (Fig. 3c), confirming that the modification is indeed ubiquitination. In parallel experiments using Flag-tagged β -TrCP and Skp2, respectively, anti-Flag immunoprecipitates were incapable of efficiently ubiquitinating either phosphorylated or dephosphorylated cyclin E even though SCF complexes were formed (Fig. 3a, b). In contrast, immunoprecipitated SCF^{Skp2} could ubiquitinate phosphorylated p27^{Kip1}, one of its established targets (Fig. 3d). Consistent with the binding studies described above, SCF^{hCdc4}-mediated ubiquitination of cyclin E (T62A) was slightly reduced compared with wild-type cyclin E, ubiquitination of cyclin E (T380A) was moderately reduced, and the double mutant was not ubiquitinated at all (Fig. 3a). Thus hCdc4 is incorporated into an SCF complex that efficiently ubiquitinates phosphorylated but not unphosphorylated cyclin E. These data, taken together with the ability of transduced wild-type and dominant negative hCdc4 to affect dramatically the steady-state levels of cyclin E *in vivo*, strongly suggest that SCF^{hCdc4} represents the predominant pathway mediating turnover of cyclin E in mammalian cells.

A tissue RNA blot indicated that hCdc4 is expressed in most, if not all, tissues. In most tissues, the predominant messenger RNA species is about 5.5 kilobases (kb) long (data not shown). However, some tissues strongly expressed a 4-kb mRNA as the predominant form, in particular brain and skeletal muscle. The fact that hCdc4 is expressed at high levels in non-proliferating tissues suggests a function in addition to turnover of cyclin E, because cyclin E expression should be limited to tissues undergoing cell division. These data also suggest that the 5.5-kb mRNA, which is ubiquitously expressed, encodes the F-box protein responsible for targeting cyclin E. Using exon-specific probes, we have shown that the 5.5-kb mRNA encodes the 110K hCdc4 isoform, whereas the 4-kb species encodes the 69K species (data not shown). Analysis of hCdc4 expression in synchronized HeLa cells indicated that neither hCdc4 mRNA nor protein is regulated during the cell cycle (data not shown).

Levels of cyclin E are elevated in many types of human

malignancy^{5,6}. Furthermore, dysregulation of cyclin E levels has been directly linked to genomic instability⁷ and tumorigenesis in model systems¹⁸. To determine whether loss of hCdc4 might account for elevated levels of cyclin E, we first analysed a panel of cell lines derived from breast cancer for cyclin E levels (Fig. 4a). Two such cell lines (MDA-MB-157 and SUM149PT, lanes 6 and 7, respectively) exhibited significant elevation of cyclin E above the level observed in 184A1 (ref. 19), an immortalized, non-transformed breast epithelial cell line (Fig. 4a, lane N). One of the cell lines expressing high levels of cyclin E (MDA-MB-157) has been shown previously to contain a genomic amplification of the cyclin E locus²⁰. Northern blot analysis using the hCdc4 cDNA as a probe (Fig. 4a) indicated that the 184A1 breast epithelial cell line contained a predominant hybridizing mRNA species at approximately 5.5 kb and a species with lower abundance at about 4 kb. Most of the breast cancer cell lines expressed only the 5.5-kb species. However, one cell line (SUM149PT) that exhibited high levels of cyclin E expressed an mRNA species of reduced mobility. The lack of any of the mRNA species characteristic of hCdc4 suggests a mutational lesion and, furthermore, loss of heterozygosity.

To determine whether the aberrant mRNA species in the SUM149PT cell line corresponded to a loss or alteration of hCdc4 protein, hCdc4 was concentrated from lysates either by adsorption to immobilized phosphorylated cyclin E or by immunoprecipitation with anti-hCdc4 antibody. The concentrated hCdc4 was then, in each case, subjected to SDS–PAGE and western blotting with anti-hCdc4 antibody (Fig. 4b). The 184A1 cell line and a breast cancer cell line that does not exhibit elevated cyclin E levels (T47D) contained the 110K hCdc4 isoform, which was detected either by binding specifically to phosphorylated cyclin E or by immunoprecipitation with anti-hCdc4 antibody. However, SUM149PT expressed no hCdc4-crossreactive protein capable of being immunoprecipitated or that bound to phosphorylated cyclin E within the limit of detection. 35 S-methionine pulse-chase experiments support this interpretation in that cyclin E has an extended half-life in the SUM149PT cell line compared with 184A1 and T47D cell lines (Fig. 4c).

To determine the nature of the mutation at the *hCDC4* locus in the SUM149PT cell line, the presumptive protein-coding region of the cDNA was sequenced and found to contain a direct repeat of exons 8 and 9 separated by 11 base pairs of intronic sequence (Fig. 4d). This mutation would be predicted to result in chain termination, eliminating the last four (of seven) WD40 repeats, presumably rendering the resulting polypeptide nonfunctional. Indeed, translation *in vitro* of the cDNA isolated from the SUM149PT cell line produced a truncated product that did not bind to phosphorylated cyclin E (data not shown). The loss of heterozygosity and internal genomic duplication at the *hCDC4* locus was confirmed by Southern blotting (Fig. 4e). DNA was cleaved with *Sst*I, which cuts in intronic sequences immediately downstream of exon 7, and with *Eco*RV, which cuts in intronic sequences immediately downstream of exon 10, and probed with a genomic fragment containing exons 8 and 9. The predicted wild-type fragment is 5.5 kb long, whereas that of the mutant is close to 11 kb. This finding and the implication of elevated cyclin E in carcinogenesis suggests that hCdc4 may be a tumour suppressor associated with some types of malignancy, including breast cancer. hCdc4 was identified independently on the basis of homology to the *Drosophila archipelago* gene product, also shown to regulate cyclin E proteolysis, and given the name human Ago in that study²¹. □

Methods

Plasmids and baculovirus constructions

A human EST encoding part of the hCdc4 gene was amplified from HeLa mRNA by polymerase chain reaction with reverse transcription (RT-PCR) using two sequence-

specific oligonucleotide primers, Pcr1 (5'-gcaagcttatgggtttctacggcaccat-3', forward), and Pcr2 (5'-atggccctgctcttctacatgtcc-3', reverse), and TA cloned into pCR2.1 (Invitrogen). The sequence of the cloned cDNA was verified in its entire length (1.7 kb) by sequencing and found to match the sequence published in the NCBI database (<http://www.ncbi.nlm.nih.gov/Genbank>, Genbank accession number BAA91986.1). For further details, see Supplementary Information. A mammalian transfection plasmid expressing N-terminal Flag-tagged hCdc4 protein was constructed by subcloning into pFLAG-CMV2 (Sigma). For expression in *Escherichia coli*, hCdc4 was tagged at the N terminus with a RGS.His epitope through subcloning into pQE-10 (Qiagen). Complementary DNAs encoding hCdc4 and a Δ F-box mutant that had been deleted for its F-box by a two-step PCR protocol, as well as the cDNA coding for β -galactosidase, were cloned into pDV46.

Recombinant adenoviruses were generated by co-transfecting the recombinant plasmids and pBHG10 (ref. 22) into 293 cells using the calcium phosphate precipitation method. The β -TrCP and Skp2 clones were gifts of F. Mercurio and M. Pagano, respectively, and were cloned into pFLAG-CMV2 to obtain β -TrCP and Skp2 tagged at their N termini with the Flag epitope. The mammalian transfection plasmid pCDNA3-Cul1-HA was a gift of R. Klausner, and pCDNA3-3MYCROC1 as well as pCDNA3-hSkp1 were gifts of Y. Xiong. Baculovirus expressing cyclin E with a GST tag at its N terminus was a gift of B. Sarcevic. Recombinant baculoviruses expressing GST-tagged versions of cyclin E phosphorylation site mutants (T62A, T380A, T62A/T380A) were generated using the pFastBac-system (Gibco BRL) according to the manufacturer's protocol. Baculovirus encoded proteins were expressed in SF9 insect cells grown in Ex-Cell 401 media (JRH) supplemented with 2% fetal bovine serum.

Analysis of cyclin E turnover in yeast

All yeast strains are isogenic to 15Daub Δ , a *bar1 Δ ura3 Δ ns*, a derivative of BF264-15D (ref. 23). Several thermosensitive *skp1* mutants with different cell cycle arrest phenotypes were constructed by a combination of PCR mutagenesis and *in vivo* gap repair similar to the procedure described by Muhlrads *et al.*²⁴. The mutant shown in Fig. 1 (*skp1-24*) arrested with 1C DNA content and a multi-budded phenotype. To analyse turnover of cyclin E in various yeast mutants expressing cyclin E from the inducible *GAL1* promoter, cells were grown in YEP-rafinosose at 25°C to an absorbance at wavelength 600 nm of 0.3. Cells were then shifted to 35°C and after 30 min galactose was added to a final concentration of 2% to induce the *GAL1* promoter. To terminate cyclin E expression after 60 min, cells were collected on filters and transferred to YEPD media and incubation was continued at 35°C. Extracts were prepared from aliquots taken after the periods indicated and analysed for cyclin E by western blot. For further details, see Supplementary Information.

Cell culture and immunological techniques

A panel of cell lines derived from breast cancer was obtained from the American Type Culture Collection (ATCC) and the University of Michigan Breast Cell/Tissue Bank and Database; the cells were grown in media recommended by the suppliers. HeLa, KB (human epidermoid carcinoma) and 293T cells were grown in DMEM (Gibco BRL) supplemented with 10% fetal bovine serum. All cells were maintained in a humidified 37°C incubator with 5% CO₂. 293T cells were transfected with various combinations of plasmids in 10-cm dishes by the calcium phosphate precipitation method. Forty hours after transfection, cells were lysed and subjected to immunoprecipitation followed by immunoblotting. For further details, see Supplementary Information.

Pulse-chase, northern and Southern blot analyses

Pulse-chase experiments were performed on thymidine-arrested cells as described²⁵. Thymidine-arrested KB cells were co-transduced with adenovirus encoding cyclin E (resulting in a five- to tenfold elevation over endogenous levels) and virus encoding either wild-type hCdc4, a Δ F-box hCdc4, or control β -galactosidase. Viral transductions were incubated for 24 h before pulse-chase. Immunoprecipitations were performed with a monoclonal anti-cyclin E antibody (HE172). Quantification of bands was performed with ImageQuant software (Molecular Dynamics). For northern blot analysis, 2 μ g of poly(A)⁺ RNA was isolated from asynchronously growing cultures according to the manufacturer's protocol (Qiagen) and run on a 1% formaldehyde agarose gel as described²⁶. The gel was blotted onto Zeta-Probe GT genomic membrane (Bio-Rad) and hybridized with a radiolabelled hCdc4 probe followed by autoradiography. For Southern blot analysis, 10 μ g of DNA was digested with *Sst*I and *Eco*RV, run on an 0.8% agarose gel, blotted and probed with a genomic fragment corresponding to exons 8 and 9.

In vitro binding

Complementary DNAs encoding various hCdc4 isoforms and mutants, β -TrCP and Skp2 were translated *in vitro* into ³⁵S-methionine-labelled proteins by a T7 transcription/translation system (Promega). GST-tagged cyclin E and various cyclin E phosphorylation-site mutants were expressed in baculovirus-infected SF9 insect cells and adsorbed on glutathione beads. Bound proteins were analysed by SDS-PAGE followed by autoradiography. For further details, see Supplementary Information.

In vitro ubiquitination assay

Recombinant SCF complexes containing different Flag-tagged F-box proteins were isolated from transfected 293T cells. Equal amounts of SCF immune complexes were mixed with cyclin E protein for 30 min on ice to allow binding. Aliquots of this mixture

were then added to ubiquitination reactions in a total volume of 30 μ l containing 15 μ g of bovine ubiquitin (Sigma), 0.5 μ g of yeast E1 enzyme (Boston Biochem), 1 μ g of human 6xHis-Cdc34 purified from bacteria, and an ATP-regenerating system (1 mM ATP, 20 mM creatine phosphate, 0.1 mg ml⁻¹ creatine kinase) in ubiquitination reaction buffer²⁵ supplemented with 5 mM sodium fluoride, 1 mM sodium orthovanadate, 1 mM phenylmethyl sulphonylfluoride, 2 μ g ml⁻¹ aprotinin, 2 μ g ml⁻¹ leupeptin and 2 μ g ml⁻¹ pepstatin. The Cdk2-inhibitor roscovitine (Biomol; 100 μ M final concentration) was added to reactions containing dephosphorylated cyclin E as substrate. Reactions were incubated at 30°C for 2 h, terminated by boiling for 5 min with SDS sample buffer, and analysed by SDS-PAGE followed by immunoblotting using anti-cyclin E antibodies. The ubiquitination assay using p27 as a substrate was performed as described²⁵.

Received 17 April; accepted 28 August 2001.

- Reed, S. I. Cyclin E: in mid-cycle. *Biochim. Biophys. Acta* **1287**, 151–153 (1996).
- Reed, S. I. Control of the G1/S transition. *Cancer Surv.* **29**, 7–23 (1997).
- Sauer, K. & Lehner, C. F. The role of cyclin E in the regulation of entry into S phase. *Prog. Cell Cycle Res.* **1**, 125–139 (1995).
- Eklholm, S. V. & Reed, S. I. Regulation of G₁ cyclin-dependent kinases in the mammalian cell cycle. *Curr. Opin. Cell Biol.* **12**, 676–684 (2000).
- Donnellan, R. & Chetty, R. Cyclin E in human cancers. *FASEB J.* **13**, 773–780 (1999).
- Sandhu, C. & Slingerland, J. Deregulation of the cell cycle in cancer. *Cancer Detect. Prev.* **24**, 107–118 (2000).
- Spruck, C. H., Won, K. A. & Reed, S. I. Deregulated cyclin E induces chromosome instability. *Nature* **401**, 297–300 (1999).
- Won, K. A. & Reed, S. I. Activation of cyclin E/CDK2 is coupled to site-specific autophosphorylation and ubiquitin-dependent degradation of cyclin E. *EMBO J.* **15**, 4182–4193 (1996).
- Clurman, B. E., Sheaff, R. J., Thress, K., Groudine, M. & Roberts, J. M. Turnover of cyclin E by the ubiquitin-proteasome pathway is regulated by cdk2 binding and cyclin phosphorylation. *Genes Dev.* **10**, 1979–1990 (1996).
- Tyers, M. & Jorgensen, P. Proteolysis and the cell cycle: with this RING I do thee destroy. *Curr. Opin. Genet. Dev.* **10**, 54–64 (2000).
- Deshais, R. J. SCF and Cullin/Ring H2-based ubiquitin ligases. *Annu. Rev. Cell Dev. Biol.* **15**, 435–467 (1999).
- Koepp, D. M., Harper, J. W. & Elledge, S. J. How the cyclin became a cyclin: Regulated proteolysis in the cell cycle. *Cell* **97**, 431–434 (1999).
- Patton, E. E., Willems, A. R. & Tyers, M. Combinatorial control in ubiquitin-dependent proteolysis: don't Skp the F-box hypothesis. *Trends Genet.* **14**, 236–243 (1998).
- Dealy, M. J. *et al.* Loss of *Cul1* results in early embryonic lethality and dysregulation of cyclin E. *Nature Genet.* **23**, 245–248 (1999).
- Wang, Y. *et al.* Deletion of the *Cul1* gene in mice causes arrest in early embryogenesis and accumulation of cyclin E. *Curr. Biol.* **9**, 1191–1194 (1999).
- Yeh, K. H. *et al.* The F-box protein Skp2 binds to the phosphorylated threonine 380 in cyclin E and regulates ubiquitin-dependent degradation of cyclin E. *Biochem. Biophys. Res. Commun.* **281**, 884–890 (2001).
- Nakayama, K. *et al.* Targeted disruption of Skp2 results in accumulation of cyclin E and p27^{Kip1}, polyploidy and centrosome overduplication. *EMBO J.* **19**, 2069–2081 (2000).
- Bortner, D. M. & Rosenberg, M. P. Induction of mammary gland hyperplasia and carcinomas in transgenic mice expressing human cyclin E. *Mol. Cell. Biol.* **17**, 453–459 (1997).
- Walen, K. H. & Stampfer, M. R. Chromosome analyses of human mammary epithelial cells at stages of chemical-induced transformation progression to immortality. *Cancer Genet. Cytogenet.* **37**, 249–261 (1989).
- Keyomarsi, K. & Pardee, A. B. Redundant cyclin overexpression and gene amplification in breast cancer cells. *Proc. Natl Acad. Sci. USA* **90**, 1112–1116 (1993).
- Moberg, K. H., Bell, D. W., Wahner, D. C. R., Haber, D. A. & Hariharan, I. K. Archipelago regulates Cyclin E levels in *Drosophila* and is mutated in human cancer cell lines. *Nature* **413**, 311–316 (2001).
- Bett, A. J., Haddara, W., Prevec, L. & Graham, F. L. An efficient and flexible system for construction of adenovirus vectors with insertions or deletions in early regions 1 and 3. *Proc. Natl Acad. Sci. USA* **91**, 8802–8806 (1994).
- Reed, S. I., Hadwiger, J. A. & Lorincz, A. T. Protein kinase activity associated with the product of the yeast cell division cycle gene CDC28. *Proc. Natl Acad. Sci. USA* **82**, 4055–4059 (1985).
- Muhlrads, D., Hunter, R. & Parker, R. A rapid method for localized mutagenesis of yeast genes. *Yeast* **8**, 79–82 (1992).
- Spruck, C. H. *et al.* A CDK-independent function of mammalian Cks1: Targeting of SCF^{Skp2} to the CDK inhibitor p27^{Kip1}. *Mol. Cell* **7**, 639–650 (2001).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, 1989).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank R. Klausner, F. Mercurio, M. Pagano, B. Sarcevic and Y. Xiong for providing reagents and D. von Seggern for help with adenovirus work. H.S. acknowledges support from the Breast Cancer Research Program (BCRP) of the University of California, and from the Breast Cancer Research Program of the US Department of Defense (DOD). C.S. is supported by a Special Fellow Award from the Leukemia and Lymphoma Society of America. P.K. is an APART fellow of the Austrian Academy of Sciences. This work was supported by a grant from the National Cancer Institute to S.I.R.

Correspondence and requests for materials should be addressed to S.I.R. (e-mail: sreed@scripps.edu). The GenBank accession number for the hCdc4 cDNA encoding the isoform with M_r 110K is AY049984.

A mouse knock-in model exposes sequential proteolytic pathways that regulate p27^{Kip1} in G1 and S phase

Nisar P. Malek^{*†‡}, Holly Sundberg^{†‡}, Seth McGrew^{*†}, Keiko Nakayama[§], Themis R. Kyriakidis^{||} & James M. Roberts^{*†||}

^{*} Howard Hughes Medical Institute and [†] Department of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98104, USA

[§] Laboratory of Embryonic and Genetic Engineering, Medical Institute of Bioregulation, Kyushu University, Fukuoka 812-8582, Japan

^{||} Department of Biochemistry, University of Washington, Seattle, Washington 98104, USA

The protein p27^{Kip1} is an inhibitor of cell division¹. An increase in p27 causes proliferating cells to exit from the cell cycle, and a decrease in p27 is necessary for quiescent cells to resume division^{2,3}. Abnormally low amounts of p27 are associated with pathological states of excessive cell proliferation, especially cancers^{4–8}. In normal and tumour cells, p27 is regulated primarily at the level of translation^{9–11} and protein turnover. Phosphorylation of p27 on threonine 187 (T187) by cyclin-dependent kinase 2 (Cdk2) is thought to initiate the major pathway for p27 proteolysis^{12–15}. To critically test the importance of this pathway *in vivo*, we replaced the murine p27 gene with one that encoded alanine instead of threonine at position 187 (p27^{T187A}). Here we show that cells expressing p27^{T187A} were unable to downregulate p27 during the S and G2 phases of the cell cycle, but that this had a surprisingly modest effect on cell proliferation both *in vitro* and *in vivo*. Our efforts to explain this unexpected result led to the discovery of a second proteolytic pathway for controlling p27, one that is activated by mitogens and degrades p27 exclusively during G1.

Phosphorylation of p27 on T187 by Cdk2 creates a binding site for a Skp2-containing E3 ubiquitin-protein ligase known as SCF^{16–18}; ubiquitination of p27 by SCF results in degradation of p27 by the proteasome^{19–22}. We tested the role of T187 phosphorylation in determining abundance of p27 protein and in controlling cell proliferation, by constructing a mouse strain that expresses a non-phosphorylatable form of p27 (p27^{T187A}) instead of the wild-type p27 protein. Control experiments showed that p27^{T187A} and wild-type p27 were equally able to inhibit cyclin E–Cdk2 when either histone 1 (H1) or the retinoblastoma (Rb) protein was used as a substrate *in vitro*¹². Therefore, we attributed the effects of this mutation, as described below, to altered p27 regulation rather than to an intrinsic change in its molecular properties as a Cdk inhibitor. In brief, the p27 gene was precisely replaced with an allele in which the codon for T187 was changed to one encoding A187. Previous experiments using cultured fibroblasts had shown that ectopic overexpression of p27^{T187A} imposed an irreversible G1 arrest¹². Therefore, a 'lox-STOP-lox' element was placed within the p27^{T187A} promoter so that the mutant allele could be conditionally activated with Cre recombinase (see Methods) (Fig. 1a). Mice heterozygous for p27^{T187A} (lox-STOP-lox) (see Methods) were bred to mice that constitutively express Cre recombinase in the germ line, which resulted in deletion of the lox-STOP-lox element. The p27^{T187A} (Δlox-STOP-lox) allele was bred to homozygosity and shown to express the p27^{T187A} protein at levels equivalent to the wild-type allele in all tissues (data not shown). Surprisingly, expression of p27^{T187A} did not affect viability or fertility. Consequently, all

further experiments were performed on mice homozygous for the p27^{T187A} (Δlox-STOP-lox) allele and that did not contain the Cre recombinase transgene.

The effect of the T187A mutation on regulation of the p27 protein was determined by comparing the protein levels of p27 and p27^{T187A} in mouse embryonic fibroblasts (MEFs). The MEFs were made quiescent by serum deprivation and then stimulated to enter the cell cycle by re-addition of serum. In control MEFs, p27 protein declined to a low level 12–15 h after serum stimulation, a time that corresponded to the early/mid-G1 phase of the cell cycle, and it remained at a low level for the duration of the cell cycle (Fig. 1b). p27^{T187A} was expressed at the same level as wild-type p27 in quiescent MEFs, a result consistent with our observation that p27 and p27^{T187A} were expressed at equal levels in mouse tissues, which are composed largely of non-dividing cells. Serum stimulation caused the level of the p27^{T187A} protein to decline with kinetics similar to wild-type p27. This was consistent with the fact that decreased expression of p27 in serum-stimulated fibroblasts preceded activation of cyclin E–Cdk2 (ref. 9 and see also Supplementary Information Fig. 1). In contrast to the wild-type protein, however, p27^{T187A} re-accumulated as cells completed G1 and entered S phase. Indeed, in late S/G2 the amount of p27^{T187A} rose to a level that was similar to its abundance in quiescent cells. This was associated with an increased amount of p27 bound to cyclin A

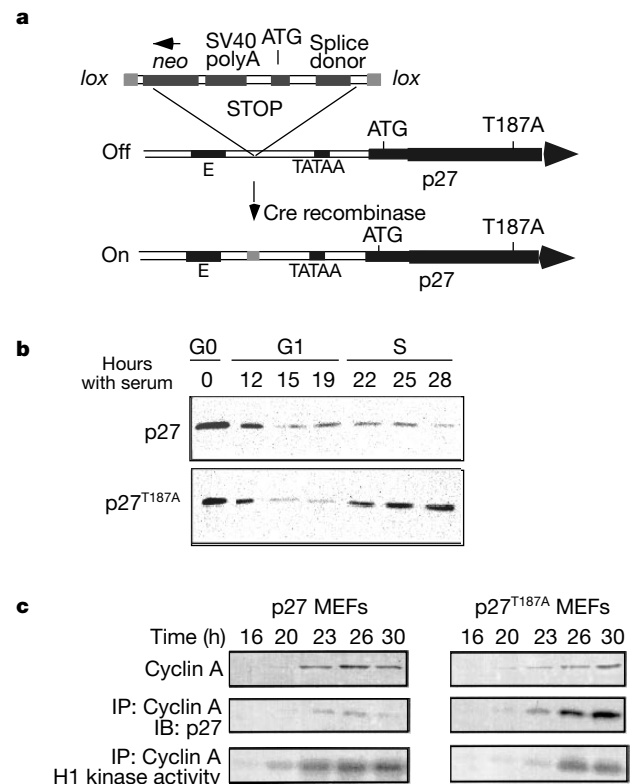


Figure 1 Replacement of the p27 gene with one encoding p27^{T187A}. **a**, Schematic diagram of the murine p27 gene modified to inducibly express a mutated p27 protein, p27^{T187A}. **b**, Immunoblot of wild-type p27 (top) and p27^{T187A} protein in serum-starved and re-stimulated MEFs. Analysis was performed using five independently isolated MEF strains from three different founder mice, and a representative result is shown. Cell cycle position at each point was determined by flow cytometry of nuclear DNA content, and by incorporation of BrdU into chromosomal DNA. **c**, MEFs expressing wild-type p27 or p27^{T187A} were arrested in G0 by serum starvation. Re-addition of serum caused synchronous re-entry into the cell cycle. Expression of cyclin A, cyclin A-associated p27, and cyclin A-associated H1 kinase activity were measured at each point. IB, western immunoblot; IP, immunoprecipitate.

[†] Present address: Department of Gastroenterology and Hepatology, and Institute for Molecular Biology, Hannover Medical School, Hannover 30625, Germany.

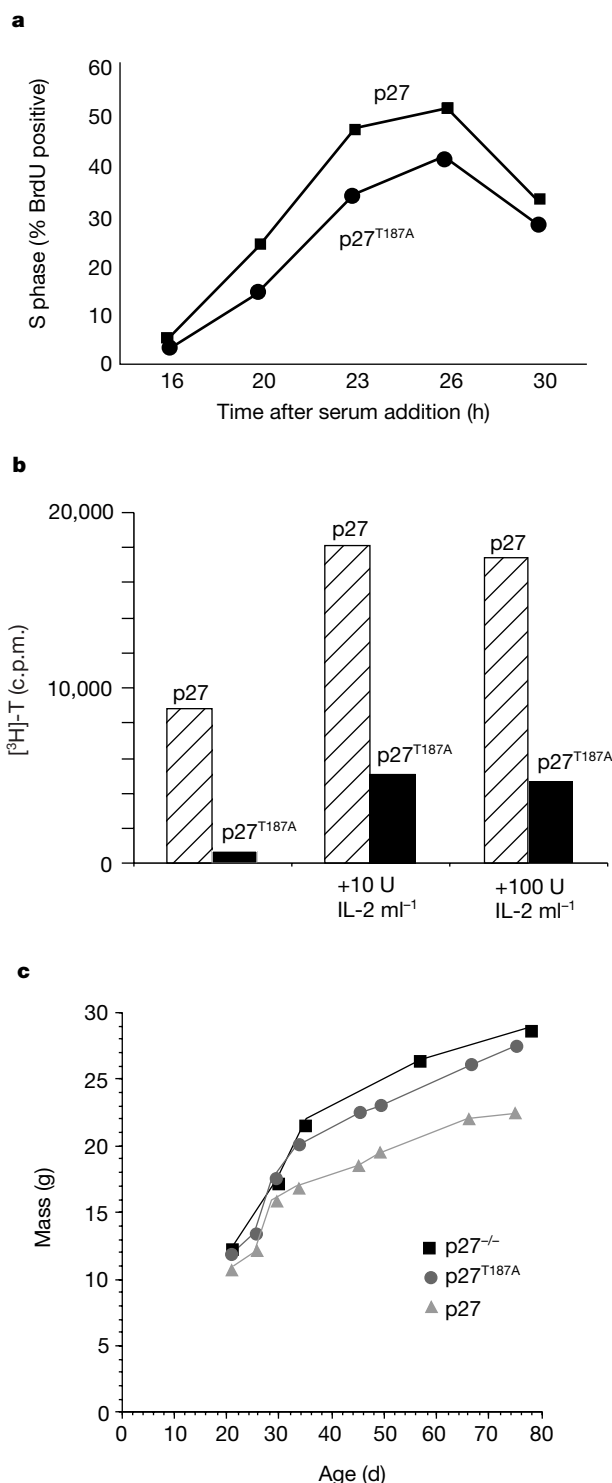


Figure 2 Effect of p27^{T187A} on cell proliferation. **a**, Quiescent MEFs were stimulated to proliferate by re-feeding with serum. Percentage of cells in S phase was determined by pulse labelling with BrdU. The analysis was repeated using three independently isolated MEF strains from three different founder mice. A representative result is shown. **b**, Purified CD4⁺ splenic T lymphocytes expressing either p27 or p27^{T187A} were stimulated to proliferate with antibodies against the T-cell antigen receptor alone, or with anti-receptor antibodies plus the indicated amount of recombinant IL-2. Proliferation was measured by incorporation of [³H]-thymidine ([³H]-T) into DNA. All measurements were done in triplicate, using lymphocytes isolated from two different founder mice. **c**, Growth curves of female wild-type p27, p27-null and homozygous p27^{T187A} mice. Each point represents an average of 30 mice. The p27 and p27^{T187A} mice were littermates (F₂ hybrids, B6/C57 × 129/Sv). Mass data for the p27-null mice were obtained from an earlier study, which used mice of the same genetic background as those used here².

and a 50% reduction in cyclin A-associated kinase activity (Fig. 1c). There was no change in total levels of cyclin A protein, and the length of S phase was not altered. These experiments demonstrated that p27 was downregulated in a T187-dependent manner in S and G₂, and independently of T187 (and Cdk2) in G₁.

The absence of the T187-dependent pathway for p27 turnover had significant effects on cell proliferation in various cells and tissues of the p27^{T187A} mutant mouse. In general we found that the rising levels of p27^{T187A} that occurred in late G₁/S/G₂ cells created a barrier to cell cycle progression. However, the severity of the ensuing proliferation defect varied among cell types. A modest effect was seen in MEFs, where expression of p27^{T187A} caused a 20–30% reduction in the number of cells that entered S phase after serum stimulation (Fig. 2a). A relatively greater defect was seen when purified splenic T lymphocytes that carry the CD4 antigen were stimulated to proliferate with antibodies directed against the T-cell antigen receptor. DNA replication was reduced by 80% in cells expressing p27^{T187A} compared with control T cells (Fig. 2b). Addition of exogenous interleukin-2 (IL-2) partially restored proliferation of the cells expressing p27^{T187A}, suggesting that high levels of IL-2 might promote a T187-independent pathway for decreasing p27 levels.

A defect in cell proliferation was also observed in dermal keratinocytes expressing p27^{T187A}. Keratinocyte proliferation was induced *in vivo* by creating circular (4 mm diameter) punch wounds of full thickness (extending through the epidermis and dermis) in the skin overlying the scapula. The rate of healing was monitored by gross inspection and by histological examination at 4.5 days after wounding. Wound closure by regrowth of epithelial cells was delayed in the p27^{T187A} mice. The epithelial gap (measured as the distance between the keratinocyte edges growing into the wound bed) made up 60 ± 5% of the entire wound in the p27^{T187A} mice as compared with 35 ± 9% in the control mice (*n* = 12) (see Supplementary Information Fig. 2). This difference was most probably a result of an impaired proliferative response, as p27^{T187A} keratinocytes at the wound edge displayed reduced incorporation of bromodeoxyuridine (BrdU) (p27^{T187A} 13.5 ± 8.5% versus control 35 ± 5%). No difference was observed between p27^{T187A} and control mice in the healing of incision wounds, which occurs mostly by epithelial cell migration rather than by proliferation (data not shown).

It was remarkable that, despite the restraint on cell proliferation created by the p27^{T187A} mutation, mice expressing this protein developed normally and attained an average size that was even larger than wild-type mice (Fig. 2c). The mechanism underlying this is unknown, but it was important to rule out the possibility that the larger size of the mutant mice was due to increased cell proliferation, because this would be inconsistent with the proliferative impairment of p27^{T187A}-expressing cells, as described above. To address this issue we examined cell proliferation in the thymus of the p27^{T187A} mice, which, like all other organs, was enlarged in proportion to overall body size. In contrast to what is observed in p27-knockout mice⁵, thymocytes expressing p27^{T187A} did not show an increased amount of cell proliferation as determined by BrdU labelling (data not shown). This indicates that the T187A mutation and the p27 gene deletion affected organ size by different mechanisms. Other phenotypes associated with p27 deficiency were also absent in the mice expressing p27^{T187A}, including female sterility, pituitary tumorigenesis and disrupted retinal architecture. Hence, there is no evidence that the p27^{T187A} mutation results in a hypoactive allele. We are investigating the effect of the p27^{T187A} mutation on apoptosis and cell size to determine whether these might contribute to increased growth rates and increased body size. However, staining kidneys, spleens and livers from 2-month-old mice with TUNEL (TdT-mediated dUTP nick end labelling) did not reveal any differences between control mice and mice expressing p27^{T187A}, and flow cytometric measurements did not show increased

size of p27^{T187A} thymocytes from 8-week-old mice (see Supplementary Information Fig. 3).

The above experiments showed that p27 abundance is controlled by two mechanisms, the first acting in early/mid-G1 cells and the second in late G1, S and G2. We made the unexpected observation that increased turnover of the p27 protein contributed not only to the T187 pathway for p27 regulation, but to the earlier G1 pathway as well (Fig. 3a). In quiescent cells, p27 was relatively stable, with a half-life of 10–12 h. Serum stimulation decreased p27 stability, reducing its half-life to about 2 h in both G1- and S-phase cells. p27^{T187A} was also stable in quiescent cells, and after serum stimulation became unstable in mid-G1 similarly to the wild-type protein. An independent method for measuring protein half-lives (pulse-chase) confirmed that wild-type p27 and p27^{T187A} had equally short half-lives (<4 h) in G1-phase MEFs. In S-phase cells, however, p27^{T187A} was stable, acquiring a long half-life very similar to that in quiescent, mitogen-starved cells. Thus, rapid turnover of p27 in S-phase cells requires T187, whereas the rapid turnover of p27 in G1 cells does not. These data also implied that the proteolytic pathway that degraded p27 in G1 cells was not operative in S phase.

This G1-specific turnover pathway for p27 is not a unique feature of cells as they exit quiescence, but rather occurs during each mitotic cycle. Quiescent MEFs were stimulated with serum mitogens for 18 h, at which time they were separated by flow cytometry into G1- and S-phase populations. Levels of the p27 protein were lower in G1 cells than in quiescent cells. In the S-phase population, however, the abundance of wild-type p27 declined further, whereas the opposite occurred in MEFs expressing p27^{T187A} (Fig. 3b). The S-phase cells were then replated and allowed to progress through the division cycle until a time when 80% of the cells had entered the next G1 phase. The abundance of p27^{T187A} declined again in the second G1,

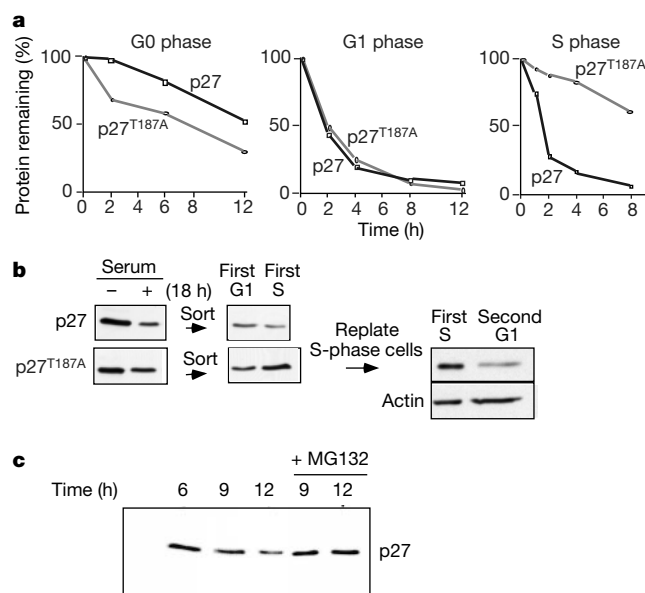


Figure 3 Cell-cycle-dependent proteolysis of p27. **a**, The half-lives of p27 and p27^{T187A} were measured in G0-, G1- and S-phase MEFs. Cells were synchronized by serum starvation (G0) and re-feeding (G1 is 12 h after re-feeding; S is 24 h after re-feeding). **b**, Levels of p27 protein were determined by immunoblotting of cell extracts from cells synchronized through two cell cycles. See text for details. **c**, Control (p27^{+/+}) MEFs were synchronized by serum starvation and then stimulated to enter the cell cycle by re-feeding with serum. One group of cells was allowed to proceed through G1, while the other group was treated with the proteasome inhibitor MG132 (10 μ M) at 6 h after serum stimulation. Levels of p27 protein were then measured by immunoblotting at the indicated time points.

just as it had in the first G1, demonstrating the periodic nature of the G1-turnover pathway.

The turnover of p27 in G1 cells was proteasome dependent. Quiescent, serum-starved MEFs were re-stimulated with serum and 6 h later treated with MG132, an inhibitor of proteasomal proteolysis. This prevented the normal decrease in p27 protein levels that occurs in mid-G1 (Fig. 3c).

Phosphorylation of p27 on T187 triggers its ubiquitination by the Skp2-containing SCF E3 complex, and its subsequent turnover in the proteasome^{12–14,19–22}. This pathway is operational in S- and G2-phase cells, after Cdk2 is activated by cyclins E and A. Surprisingly, we found that the G1-specific turnover of p27 was also mediated by Skp2 in MEFs, but independently of phosphorylation on T187. Extracts from asynchronously proliferating MEFs (not shown) and MEFs synchronized in G1 (Fig. 4a) were immunoprecipitated with anti-Skp2 antibodies. p27 and p27^{T187A} were detected in equal abundance in Skp2 complexes. The binding of Skp2 to both p27 and p27^{T187A} was phosphorylation dependent, because treatment of isolated p27–Skp2 and p27^{T187A}–Skp2 complexes with a protein phosphatase disrupted the interaction (Fig. 4b, Supplementary Information Fig. 4). The phosphorylation site(s) on p27 and/or Skp2 that mediate this interaction have not yet been identified. We further showed that Skp2 was essential for the turnover of p27 in G1 cells, because levels of p27 protein did not decline, either in G1 or in S phase, in serum-stimulated, early-passage, Skp2-null MEFs (Fig. 4c)²³. In conclusion, in MEFs the trigger for p27 turnover differs in G1 and S phase, but both pathways ultimately lead to the degradation of p27 by Skp2 and proteasome-dependent mechanisms.

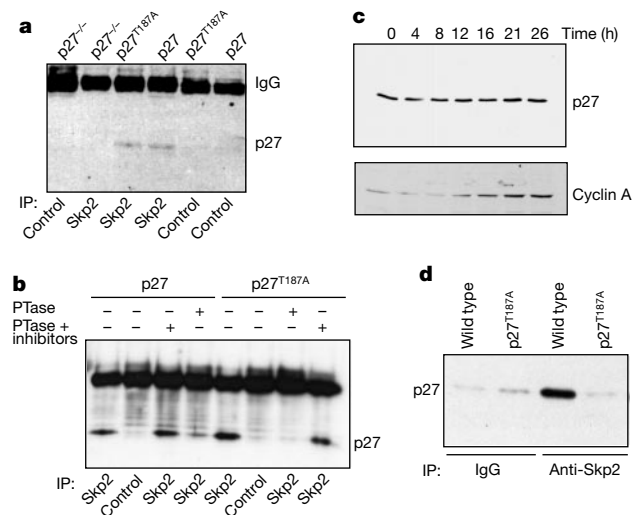


Figure 4 Turnover of p27 in G1- and S-phase MEFs requires Skp2. **a**, Skp2 was immunoprecipitated from MEFs synchronized in G1 by serum starvation and re-feeding. The presence of associated p27 (or p27^{T187A}) was determined by immunoblotting. Control immunoprecipitations used an irrelevant immunoglobulin- γ (IgG, anti-cytochrome C), and a control immunoprecipitation from p27-null MEFs is also shown. **b**, A plasmid encoding Skp2 was cotransfected into human 293T cells together with a plasmid encoding either p27 or p27^{T187A}. Whole-cell extracts were immunoprecipitated with anti-Skp2 or control (anti-cytochrome C) antibodies and, as indicated, immunoblotted directly for associated p27, treated with alkaline phosphatase (PTase) before immunoblotting, or treated with alkaline phosphatase plus phosphatase inhibitors. **c**, Quiescent Skp2^{-/-} MEFs were stimulated to re-enter the cell cycle by addition of serum, and levels of p27 protein were determined at the indicated time points by immunoblotting. Skp2^{-/-} MEFs progressed through G1 and entered S phase about 16 h after serum stimulation, as determined by the timing of cyclin A expression. **d**, The binding of p27 and p27^{T187A} to Skp2 was determined in extracts of purified thymocytes from control and homozygous p27^{T187A} mice. IP, immunoprecipitate.

In contrast to the situation in MEFs, a p27^{T187A}–Skp2 complex was not detected in purified murine thymocytes (Fig. 4d). This reflected the relatively more severe effect of the p27^{T187A} mutation on proliferation in these cells than in MEFs, and underscored the importance of a T187-independent pathway for p27 turnover in allowing for normal cell proliferation. However, it is not yet known whether the enhanced proliferation of p27^{T187A} thymocytes caused by exogenous IL-2 (Fig. 2b) is due to stimulation of the p27^{T187A}–Skp2 interaction or whether it may be through a Skp2-independent mechanism for p27 turnover.

Our data show that T187-dependent turnover of p27 is important for normal regulation of p27 and normal control of cell division. However, contrary to what might have been expected, inactivating this pathway has neither a universal nor a severe effect on cell proliferation. This may be explained, at least in part, by a previously unrecognized T187-independent pathway for p27 degradation that is activated during each G1 phase of the cell cycle. This pathway allows many cells expressing p27^{T187A} to complete the cell cycle before the re-accumulation of p27 in S phase can stop it. The T187 pathway, which keeps p27 levels low for the duration of S and G2, may allow the cell to slow its progress through this part of the cell cycle (for instance, in response to DNA damage) without experiencing continually rising and ultimately inhibitory amounts of p27.

Our major conclusion is that two proteolytic pathways act in sequence during the cell cycle to control p27 abundance. The first pathway functions during early–mid G1 and is triggered by mitogens. It may be activated by Ras and Myc, and underlie the ability of these proteins to reduce p27 abundance independently of serum stimulation^{24,25} and independently of Cdk2 (ref. 24). Inhibition of p27 at the level of translation^{9–11} and by sequestration into cyclin D–Cdk complexes²⁶ also contribute to downregulation of p27 during the early–mid-G1 cell cycle period. Downregulation of p27 by the concerted action of these pathways results in the initial production of active cyclin E–Cdk2, and consequently the onset of the second pathway for p27 turnover. This second pathway operates in late G1, S and G2, and is dependent on Cdk2-mediated phosphorylation of p27 on T187. Once initiated, this second pathway would be amplified by a positive-feedback loop, and therefore would continue even if the initial mitogenic stimulus were withdrawn. In this way, inactivation of p27 switches in mid G1 from being mitogen dependent to being mitogen independent, which is analogous to the consecutive mitogen-dependent and mitogen-independent pathways that inactivate Rb during the same cell cycle interval²⁷. Sequentially acting pathways that inactivate key cell cycle inhibitors may be the biochemical underpinnings of the cell cycle transition from mitogen dependence to mitogen independence, which has been called the G1 restriction point²⁸. □

Methods

Mice

To construct the genomic targeting vector, a 5.9-kilobase (kb) *Bam*HI fragment was isolated from a previously described 17-kb *Not*I fragment that contains the entire coding region of the p27 gene obtained from a mouse 129/SvJ genomic library². Threonine 187 of exon 2 was mutated to alanine (ACG→GCG) to create the p27^{T187A} allele. A 3-kb *Sac*I fragment containing a *pgk*-driven neomycin-resistance gene and a transcription termination cassette isolated as a *Bam*HI–*Hind*III fragment from pBS302 (Gibco BRL) was flanked with *loxP* sequences and inserted into a *Sac*I site in the p27 promoter, creating the construct p27^{T187A}(5.9-*Neo/STOP*). The p27^{T187A}(5.9-*Neo/STOP*) construct was then cloned into the pPNT vector³, thus creating the genomic targeting vector. The targeting vector was linearized with *Not*I and transduced by electroporation into mouse XY AK7 embryonic stem cells.

Embryonic stem cells were selected in 400 µg ml^{−1} G418 and 0.4 µM fialuridine. Colonies of embryonic stem cells were screened for homologous recombination events by Southern blotting using a probe external to the 5′ end of the targeting construct. In all five embryonic stem cell clones used for blastocyst injection, integration of the T187A mutation was verified by sequencing. p27^{T187A} embryonic stem cells were introduced by microinjection into C57/B6J mouse embryos at 5 d after conception. Germline transmission was identified in male chimaeras representing three separate embryonic stem cell clones. To excise the *neo/STOP* cassette from the p27 gene, chimaeric male mice were

bred with female CMV-Cre transgenic mice (TgN(CMV-Cre)1AN)²⁹. Excision of the *neo/STOP* cassette was verified by polymerase chain reaction (PCR) using primers derived from the p27^{T187A} genomic sequence upstream of the *Sac*I site (Y1, GAGCAGGTTTGTGGCAGTCGTACACCTCC), from the neomycin gene (A4, CGTGGGATCATTTGTTTCTCTCTG), and from the genomic sequence downstream of the *Sac*I site (H3, CCAATATGGCGGTGGAA GGGAGGCTGA). Homozygous integration of the T187A mutation was confirmed by the presence of a 34-base pair (bp) (*loxP* site) insertion into the wild-type 0.25-kb PCR fragment using primers Y1 and H3.

MEFs

Males and females heterozygous for p27^{T187A} were crossed and embryos were dissected 12.5–13.5 d after detection of vaginal plugs. The head and internal organs were removed and the embryos were minced and incubated in 0.05% trypsin for 5 min. The cells were resuspended in DMEM supplemented with 10% FBS. After centrifugation the supernatant was discarded and the cell suspension from each embryo was cultivated on a 10-cm dish in 8 ml DMEM with 10% FBS until confluency was reached. After this time, the cells were treated with trypsin, counted and plated at 1.4 × 10⁶ cells per 10-cm dish every 3 d.

Cell culture

In all experiments, unless otherwise noted, cell cycle position of synchronized cells was determined by flow cytometric measurement of nuclear DNA content. Passage 2–3 MEFs were plated at 1.4 × 10⁶ per 10-cm dish grown in DMEM with 10% FBS for 3 d, after which the medium was removed, the plates washed with PBS and the cells incubated in DMEM containing 0.1% FBS for 72 h. The cells were then washed with PBS, treated with trypsin, counted and resuspended in DMEM containing 10% FBS at 1.4 × 10⁶ cells per 10-cm dish and 0.5 × 10⁵ cells per 6-cm dish. For each time point, the cells were labelled with 10 µM BrdU (Sigma) for 30 min, scraped off the plate, washed with PBS and fixed in 70% ethanol for at least 24 h. Nuclei were purified and labelled with 100 µl anti-BrdU-fluorescein isothiocyanate (FITC) antibody (PharMingen). After this incubation, nuclei were treated with RNase A, counterstained with propidium iodide (100 µg ml^{−1}) and analysed on a Becton Dickinson flow cytometer using Becton Dickinson Cell Quest software.

To separate G1- from S-phase cells, cells were labelled with Hoechst Stain (Sigma) (10 µg ml^{−1}) for 30 min, treated with trypsin and separated according to DNA content on a Vantage SE flow cytometer (BD systems) using Cell Quest software. For p27 half-life measurements, cycloheximide (CHX, 10 µg ml^{−1} final concentration) was added to the cells at the indicated times. p27 protein levels were determined by immunoblotting. Cell cycle position was verified by flow cytometry and BrdU labelling. The resulting autoradiograms were scanned and the intensity of the p27 bands quantified using Image Quant software (Molecular Dynamics) and normalized to actin controls.

Isolation and stimulation of T lymphocytes

Splenic CD4⁺ T cells were purified after red-cell lysis (Whole Blood Erythrocyte Lysing Kit, R&D Systems). Mononuclear cells were enriched using Ficoll-gradient centrifugation. CD4⁺ T cells were isolated using mouse T Cell CD4 Subset Columns (R&D Systems). CD4⁺ T cells were activated with plate-bound anti-CD3 antibody with or without the addition of recombinant mouse IL-2 (PharMingen) using concentrations as indicated for the individual experiments. Seventy-two hours after stimulation, T cells were labelled with [³H]-thymidine for 4 h, collected and the incorporation of radioactivity into DNA was determined.

Immunoblotting

The antibodies used in these experiments were mouse monoclonal anti-p27 (Transduction), rabbit polyclonal anti-cyclin A (Santa Cruz), anti-actin (Santa Cruz), anti-Skp2 (Santa Cruz) and anti-Cdk2 (Santa Cruz). For immunoprecipitation of p27–Skp2 complexes, early-passage MEFs were synchronized as described above, and cells were lysed in RIPA buffer. Cell extracts were pre-cleared in protein A sepharose (Repligen) for 1 h and incubated with anti-Skp2 antibodies (Zymed GP45 and Santa Cruz H-435) overnight. Control immunoprecipitations used anti-cytochrome C antibodies (Santa Cruz, H104). After incubation with protein A sepharose, the beads were washed with RIPA, boiled in SDS sample buffer and separated on a 12% polyacrylamide gel. p27 was detected using anti-p27 antibodies. For detection of p27–Skp2 complexes in thymocytes, thymuses were removed from 2–6-week-old mice, organs were homogenized using glass slides, cells lysed in RIPA and processed as described above. For phosphatase treatment of p27–Skp2 complexes, 293 cells were transfected with expression vectors (CS2 with Skp2, p27 and p27^{T187A}). Cells were lysed in RIPA and immunoprecipitated with anti-Skp2 antibodies (Santa Cruz). p27–Skp2 complexes bound to sepharose beads were treated with alkaline phosphatase buffer (15 ml) with or without 5 U alkaline phosphatase (30 min at 30 °C in 20 µl) (both Boehringer Mannheim), or alkaline phosphatase plus a cocktail of phosphatase inhibitors (10 mM β-glycerophosphate, 100 µM sodium vanadate, 50 mM sodium fluoride). Beads were then washed three times with RIPA, boiled in SDS sample buffer, separated on a 12% polyacrylamide gel, and immunoblotted for p27.

Wound healing

Full-thickness punch wounds (4 mm diameter) were applied. Animals were injected with BrdU (1 mg ml^{−1}, 30 µl g^{−1}) 16 h before they were killed. Wounds were excised, fixed and embedded in paraffin 4.5 d after wounding³⁰. Wounds were stained with anti-BrdU antibody (AB1, NeoMarkers); BrdU staining was visualized using the Darko Ark Kit and

counterstained with haematoxylin and eosin. Wound sizes and epithelial gap diameters were determined by optical micrometer measurements at $\times 100$ magnification.

Received 25 May; accepted 18 July 2001.

- Sherr, C. J. & Roberts, J. M. Inhibitors of mammalian G1 cyclin-dependent kinases. *Genes Dev.* **9**, 1149–1163 (1995).
- Fero, M. L. *et al.* A syndrome of multiorgan hyperplasia with features of gigantism, tumorigenesis, and female sterility in p27(Kip1)-deficient mice. *Cell* **85**, 733–744 (1996).
- Coats, S. *et al.* A new pathway for mitogen-dependent cdk2 regulation uncovered in p27(Kip1)-deficient cells. *Curr. Biol.* **9**, 163–173 (1999).
- Ophascharoensuk, V., Fero, M. L., Hughes, J., Roberts, J. M. & Shankland, S. J. The cyclin-dependent kinase inhibitor p27Kip1 safeguards against inflammatory injury. *Nature Med.* **4**, 575–580 (1998).
- Fero, M. L., Randel, E., Gurley, K. E., Roberts, J. M. & Kemp, C. J. The murine gene p27^{Kip1} is haplo-insufficient for tumour suppression. *Nature* **396**, 177–180 (1998).
- Porter, P. L. *et al.* Expression of cell-cycle regulators p27^{Kip1} and cyclin E, alone and in combination, correlate with survival in young breast cancer patients. *Nature Med.* **3**, 222–225 (1997).
- Tan, P. *et al.* The cell cycle inhibitor p27 is an independent prognostic marker in small (T1a,b) invasive breast carcinomas. *Cancer Res.* **57**, 1259–1263 (1997).
- Catzavelos, C. *et al.* Decreased levels of the cell-cycle inhibitor p27^{Kip1} protein: prognostic implications in primary breast cancer. *Nature Med.* **3**, 227–230 (1997).
- Agrawal, D. *et al.* Repression of p27^{Kip1} synthesis by platelet-derived growth factor in BALB/c 3T3 cells. *Mol. Cell. Biol.* **16**, 4327–4336 (1996).
- Hengst, L. & Reed, S. I. Translational control of p27^{Kip1} accumulation during the cell cycle. *Science* **271**, 1861–1864 (1996).
- Millard, S. S., Vidal, A., Markus, M. & Koff, A. A U-rich element in the 5' untranslated region is necessary for the translation of p27 mRNA. *Mol. Cell. Biol.* **20**, 5947–5959 (2000).
- Sheaff, R. J., Groudine, M., Gordon, M., Roberts, J. M. & Clurman, B. E. Cyclin E-CDK2 is a regulator of p27^{Kip1}. *Genes Dev.* **11**, 1464–1478 (1997).
- Vlach, J., Hennecke, S., Alvezopoulos, K., Conti, D. & Amati, B. Growth arrest by the cyclin-dependent kinase inhibitor p27^{Kip1} is abrogated by c-Myc. *EMBO J.* **15**, 6595–6604 (1996).
- Muller, D. *et al.* Cdk2-dependent phosphorylation of p27 facilitates its Myc-induced release from cyclin E/cdk2 complexes. *Oncogene* **15**, 2561–2576 (1997).
- Pagano, M. *et al.* Role of the ubiquitin–proteasome pathway in regulating abundance of the cyclin-dependent kinase inhibitor p27. *Science* **269**, 682–685 (1995).
- Feldman, R. M., Correll, C. C., Kaplan, K. B. & Deshaies, R. J. A complex of Cdc4p, Skp1p, and Cdc53p/cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p. *Cell* **91**, 221–230 (1997).
- Bai, C. *et al.* SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. *Cell* **86**, 263–274 (1996).
- Skowyra, D., Craig, K. L., Tyers, M., Elledge, S. J. & Harper, J. W. F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell* **91**, 209–219 (1997).
- Sutterluty, H. *et al.* p45^{SKP2} promotes p27^{Kip1} degradation and induces S phase in quiescent cells. *Nature Cell Biol.* **1**, 207–214 (1999).
- Rolfe, M., Chiu, M. I. & Pagano, M. The ubiquitin-mediated proteolytic pathway as a therapeutic area. *J. Mol. Med.* **75**, 5–17 (1997).
- Carrano, A. C., Eytan, E., Hershko, A. & Pagano, M. SKP2 is required for ubiquitin-mediated degradation of the CDK inhibitor p27. *Nature Cell Biol.* **1**, 193–199 (1999).
- Tsvetkov, L. M., Yeh, K. H., Lee, S. J., Sun, H. & Zhang, H. p27(Kip1) ubiquitination and degradation is regulated by the SCF(Skp2) complex through phosphorylated Thr187 in p27. *Curr. Biol.* **9**, 661–664 (1999).
- Nakayama, K. *et al.* Targeted disruption of Skp2 results in accumulation of cyclin E and p27(Kip1), polyploidy and centrosome overduplication. *EMBO J.* **19**, 2069–2081 (2000).
- Leone, G., DeGregori, J., Sears, R., Jakoi, L. & Nevins, J. R. Myc and Ras collaborate in inducing accumulation of active cyclin E/Cdk2 and E2F. *Nature* **387**, 422–426 (1997).
- O'Hagan, R. C. *et al.* Myc-enhanced expression of Cul1 promotes ubiquitin-dependent proteolysis and cell cycle progression. *Genes Dev.* **14**, 2185–2191 (2000).
- Sherr, C. J. & Roberts, J. M. CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes Dev.* **13**, 1501–1512 (1999).
- Hatakeyama, M. *et al.* The cancer cell and the cell cycle clock. *Cold Spring Harb. Symp. Quant. Biol.* **59**, 1–10 (1994).
- Pardee, A. B. A restriction point for control of normal animal cell proliferation. *Proc. Natl Acad. Sci. USA* **71**, 1286–1290 (1974).
- Nagy, A. *et al.* Dissecting the role of N-myc in development using a single targeting vector to generate a series of alleles. *Curr. Biol.* **8**, 661–664 (1998).
- Kyriakides, T. R., Tam, J. W. & Bornstein, P. Accelerated wound healing in mice with a disruption of the thrombospondin 2 gene. *J. Invest. Dermatol.* **113**, 782–787 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank B. Clurman, R. Sheaff, K.-I. Nakayama, B. Carter, M. Groudine, B. Luscher and members of the Roberts laboratory for advice, suggestions and reagents. M. Black and N. Jiang provided expert technical assistance with flow cytometry and with the generation of mutant mouse strains. We thank P. Porter for help with TUNEL staining. N.P.M. was supported by a grant from the Deutsche Forschungsgemeinschaft. T.R.K. was supported by The National Science Foundation through the University of Washington Engineered Biomaterials Engineering Research Center. H.S. was supported by a grant from the National Institutes of Health. This work was supported by a grant from the National Institutes of Health to J.M.R. and by a George M. O'Brien Kidney Research Center award. J.M.R. is an Investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to J.M.R. (e-mail: jroberts@fred.hccrc.org).

RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts

Eric J. Steinmetz^{*†}, Nicholas K. Conrad^{†‡}, David A. Brow^{*} & Jeffry L. Corden[‡]

^{*} Department of Biomolecular Chemistry, University of Wisconsin Medical School, Madison, Wisconsin 53706-1532, USA

[‡] Department of Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

[†] These authors contributed equally to this work

A eukaryotic chromosome contains many genes, each transcribed separately by RNA polymerase (pol) I, II or III. Transcription termination between genes prevents the formation of polycistronic RNAs and anti-sense RNAs, which are generally detrimental to the correct expression of genes. Terminating the transcription of protein-coding genes by pol II requires a group of proteins that also direct cleavage and polyadenylation of the messenger RNA in response to a specific sequence element, and are associated with the carboxyl-terminal domain of the largest subunit of pol II (refs 1–6). By contrast, the *cis*-acting elements and *trans*-acting factors that direct termination of non-polyadenylated transcripts made by pol II, including small nucleolar and small nuclear RNAs, are not known. Here we show that read-through transcription from yeast small nucleolar RNA and small nuclear RNA genes into adjacent genes is prevented by a *cis*-acting element that is recognized, in part, by the essential RNA-binding protein Nrd1. The RNA-binding protein Nab3, the putative RNA helicase Sen1, and the intact C-terminal domain of pol II are also required for efficient response to the element. The same proteins are required for maintaining normal levels of Nrd1 mRNA, indicating that these proteins may control elongation of a subset of mRNA transcripts.

We previously identified yeast Nrd1 as an essential nuclear RNA-binding protein that directs 3' truncation of reporter gene mRNA transcripts in response to an artificial element containing a Nrd1-binding sequence^{7,8}. Nrd1-dependent 3' truncation also requires Sen1, an essential nuclear helicase-like protein that has been implicated in biogenesis of particular small nucleolar (sno) RNAs^{9–11}. The properties of Nrd1 suggested that it might direct termination by pol II, although it is not a known component of the well-characterized pre-mRNA cleavage and polyadenylation apparatus.

Direct evidence that snoRNAs might be natural targets for Nrd1 was first obtained from expression profiling of poly(A)⁺ RNA derived from temperature-sensitive *nrd1* yeast strains as compared with wild type (see Methods). Several open reading frames (ORFs) exhibiting increased expression in the *nrd1* mutant strain are located downstream of snoRNA genes in the yeast genome. For example, increased expression of transcripts containing the *TRS31* coding region, which is located immediately downstream of the gene encoding snR13 snoRNA (Fig. 1a), is observed in *nrd1* mutant strains.

Northern blot analysis of snR13 and Trs31 transcripts from wild-type and *nrd1* mutant strains shows the accumulation in the mutant strain of a new RNA species, which is larger than the normal Trs31 mRNA, that is detected by probes complementary to either mature snR13 RNA or Trs31 mRNA (Fig. 1b). The abundance of this extended product increases rapidly and markedly when the *nrd1* mutant is shifted from permissive to restrictive temperature. The 5' end of this product coincides with the snR13 RNA 5' end, as shown

by reverse transcription of RNA from a *nrd1* mutant strain (Fig. 1c). We conclude that the increased *TRS31* ORF signal in the *nrd1* mutant is caused by extended transcripts initiating from the *SNR13* promoter that, failing to terminate properly, read through downstream sequences until they encounter the *TRS31* cleavage/polyadenylation signal. The persistence of normal-length snR13 2 hours after temperature shift is expected, as snoRNAs are highly stable. The diminished accumulation of normal-sized Trs31 mRNA coincident with the accumulation of extended snR13 products suggests that read through of snR13 transcripts results in decreased initiation at the *TRS31* promoter. This apparent "promoter occlusion"¹² by elongating pol II supports the hypothesis that Nrd1 is required for terminating *SNR13* transcription.

Four other snoRNA genes implicated by the ORF array analysis (*SNR3*, *SNR47*, *SNR71* and the *SNR190-SNR128* dicistronic transcription unit) also produce extended transcripts in a *nrd1* mutant strain, as determined by reverse transcription of total RNA with downstream primers (Fig. 1c). In all cases, a *sen1* mutant strain accumulates as much as, or more read-through transcript than the *nrd1* strain, indicating that there is also a general requirement for Sen1 protein. Several other snoRNAs, including snR33, snR39b, snR45 and snR50, also exhibit increased accumulation of 3'-extended forms in *nrd1* and *sen1* mutants. The doublet of snR45 reverse transcription products is consistent with a 5'-truncated form observed previously in a *sen1* mutant¹⁰. Thus, Nrd1-dependent 3' end formation seems to be a general property of snoRNA gene expression.

Nrd1 has an amino-terminal domain that interacts with the pol II C-terminal domain (CTD)¹³. Nrd1 also interacts with another essential RNA-binding protein, Nab3, which in turn interacts genetically with Ctk1 (ref. 14), the catalytic subunit of a kinase that phosphorylates the CTD¹⁵. We proposed previously that these components act together to regulate transcript elongation¹⁴. Northern blot analysis shows the accumulation of extended forms of snR13 in a temperature-sensitive *nab3* mutant strain (Fig. 1d). A strain expressing a truncated pol II large subunit (10 repeats of the CTD heptapeptide consensus, YSPTSPS, rather than the usual 26)

and a *ctk1* null strain similarly show modest increases in accumulation of the snR13 read-through products. These results support the involvement of each of these factors in the Nrd1-dependent 3'-end formation pathway.

SnoRNA 3'-end maturation involves exonucleolytic digestion of larger precursors by the exosome, a complex of 3'-to-5' exonucleases^{16,17}. A primary 3' end is necessary for exosome entry, and efficient processing may further require that this 3' end be non-polyadenylated¹⁸. To test whether Nrd1 function is important for the synthesis of mature snoRNAs, we performed *in vivo* pulse labelling¹¹. The accumulation of newly synthesized, mature snR13 during a brief (15-min) pulse of label in cells grown at the restrictive temperature of 37 °C is decreased roughly threefold in the *nrd1* mutant strain as compared with the wild type. Synthesis of both products of the *SNR190/SNR128* dicistronic transcription unit is similarly decreased, whereas synthesis of mature snR3 is reduced about twofold (Fig. 1e). U6 RNA, a pol III transcript, is unaffected. Thus, Nrd1 function is important for efficient synthesis of mature forms of several snoRNA species.

To show further that there is a termination defect in the *nrd1* mutant, we carried out transcription run-on analysis on the endogenous *SNR3* locus (Fig. 2). We chose this locus because, in contrast to *SNR13*, the downstream ORF is in a convergent orientation, therefore single-stranded probes for *SNR3* read-through transcripts will not hybridize with mRNA from the downstream ORF. An increase in polymerase density of about fivefold is seen in the *nrd1* strain with the 5' proximal downstream probe and an almost twofold increase is seen with the distal probe. Therefore, a significant proportion of polymerase molecules that terminate transcription downstream of the *SNR3* coding region in the wild-type strain fail to do so in the *nrd1* mutant, supporting the idea that Nrd1 functions in pol II termination.

Deletion studies mapped the downstream border of the *SNR13* 3'-end formation element to within 108 base pairs (bp) of the mature 3' end of snR13 (N.K.C. and J.L.C., unpublished data). We tested the Nrd1 dependence of the *SNR13* 3'-end formation element using a reporter gene assay^{7,19} (Fig. 3a). Inserting 3'-end-

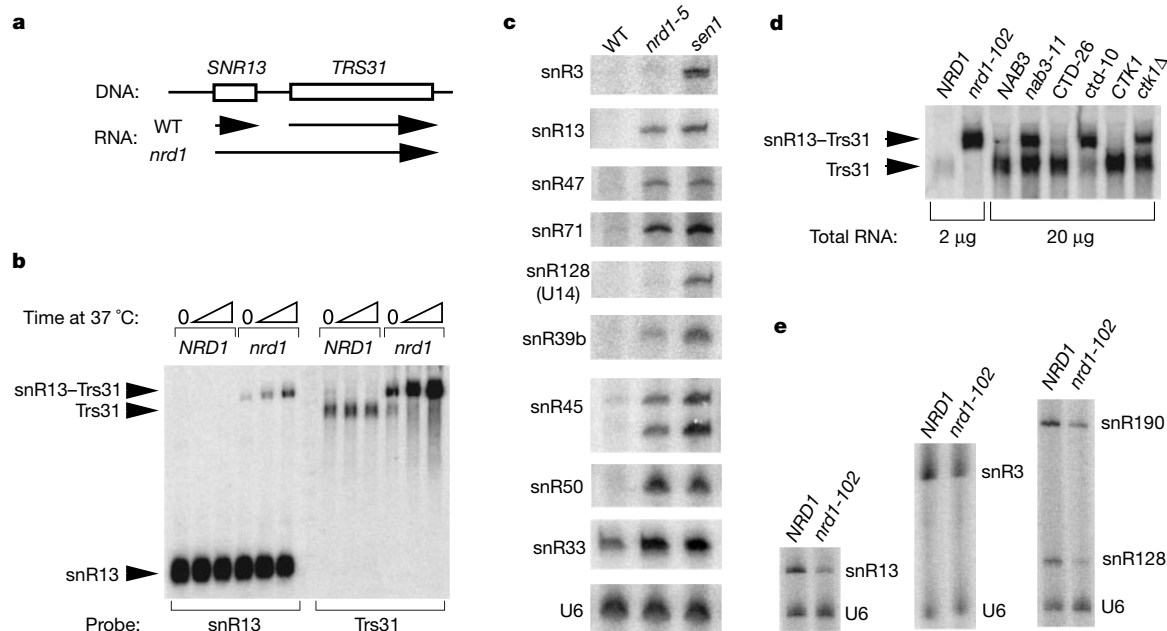


Figure 1 3'-Extended snoRNA transcripts are produced in *nrd1* and other mutant strains. **a**, Transcripts from the *SNR13-TRS31* region of yeast chromosome IV. **b**, Northern blot of snR13 and Trs31 transcripts in wild-type and temperature-sensitive *nrd1-102* mutant strains 0, 45 and 120 min after shift to 37 °C. **c**, Reverse transcription analysis of extended transcripts of several snoRNAs in *nrd1-5* and *sen1* mutant strains. **d**, Northern

blot of total RNA from *nrd1-102*, *nab3-11*, CTD truncation, *ctk1Δ* and congenic wild-type strains. Blots were probed with a *TRS31* coding region probe. **e**, Reduced rates of synthesis of specific snoRNAs in a *nrd1* mutant. Pulse-labelled RNAs were fractionated by hybrid selection and gel electrophoresis. The pol-III-transcribed U6 RNA was used as a normalization control in **c** and **e**.

forming elements into the intron of an *ACT1-CUP1* fusion gene prevents the synthesis of full-length *ACT-CUP* pre-mRNA, which means that yeast lacking the endogenous *CUP1* gene cannot grow on copper-containing media. Failure to respond to the 3'-end formation element in a *nrd1* mutant strain results in increased copper resistance. A fragment consisting of the first 108 bp of the *SNR13-TRS31* intergenic region (+125 to +232) causes severe copper sensitivity when introduced into the *ACT-CUP* intron in a wild-type strain, but copper sensitivity is relieved in the *nrd1* mutant strain (Fig. 3b). Therefore, features sufficient for Nrd1-dependent 3'-end formation reside in the 108 bp downstream of the mature 3' end of snR13, and this element is functional for 3'-end formation when transcribed from a heterologous, mRNA promoter.

Unlike most snoRNAs, each of the pol II small nuclear (sn) RNA primary transcripts contains a downstream RNA stem-loop element that specifies endonucleolytic cleavage by the Rnt1 endonuclease^{16,17,20-22}. The 3' end thus generated can provide an entry site for 3'-end maturation by the exosome. Transcription in each case is thought to continue beyond the Rnt1 cleavage site, terminating by an unknown mechanism further downstream, as extended forms of the snRNAs accumulate in *rnt1Δ* strains. For example, U4 snRNA, the 160-nucleotide (nt) product of the *SNR14* gene, is produced from larger precursors of 500–600 nt that contain Rnt1 cleavage sites at positions +295 and +329 relative to the transcription start site (Fig. 3c).

To test whether the Nrd1-dependent 3'-end formation pathway functions in biogenesis of an snRNA, portions of the *SNR14* 3'-flanking region downstream of the Rnt1 cleavage site were introduced into the *ACT-CUP* reporter gene intron. Indeed, a fragment containing most of the sequence between the Rnt1 element and the next ORF downstream (base pairs +333 to +640 relative to the *SNR14* transcription start site) causes copper sensitivity when inserted into the *ACT-CUP* intron, and this sensitivity is relieved in the *nrd1* mutant strain (Fig. 3d).

Three spontaneous mutations within the *SNR14* Nrd1-responsive region of the *ACT-CUP* reporter confer increased copper resistance. Each mutation alters a tetranucleotide located at positions +386 to +389 (Fig. 3e). As shown by the alignment in Fig. 3e, these mutations fall in a region of the *SNR14* 3' flank that has 23 nucleotide matches with a recently described 33-nt *SNR13* 3'-end formation element¹⁸. The alignment includes an 8-nt stretch of

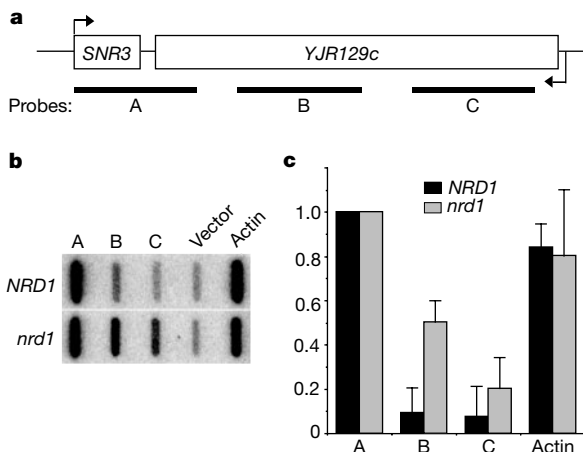


Figure 2 Termination of *SNR3* transcription is impaired in a *nrd1* strain. **a**, The *SNR3-YJR129C* genomic locus. Probes (~360 nt each) are shown below their locations. Note that *SNR3* and *YJR129C* are transcribed from opposite strands of the DNA, as indicated by the arrows. **b**, Representative slot blot from transcription run-on analysis. Vector slots are single-stranded phagemids with no insert. **c**, Quantification of the transcripts by phosphorimager analysis. Background (vector) was subtracted from each value and corrected counts were normalized to the 'A' probe. Error bars represent s.d. ($n = 4$).

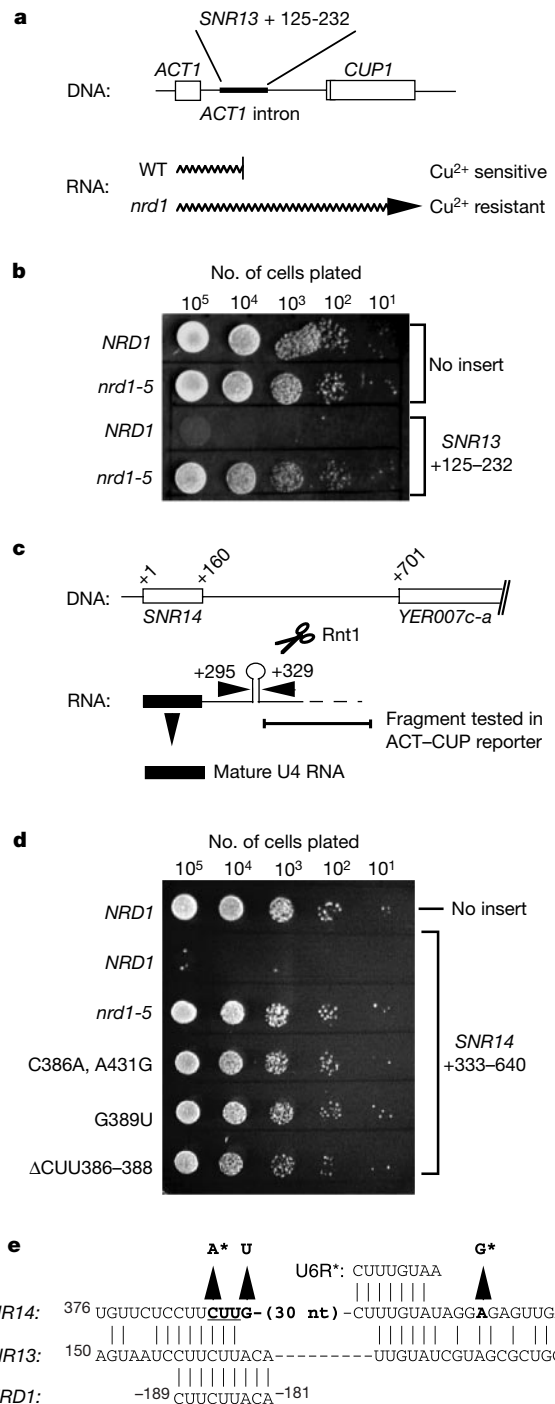


Figure 3 Nrd1-dependent 3'-end formation elements from *SNR13* and *SNR14* (U4 snRNA) genes. **a**, *ACT1-CUP1* fusion reporter gene containing *SNR13* 3'-flanking sequence. **b**, Nrd1-dependence of *SNR13* 3'-end formation element. Cells were incubated at 30 °C for 3 days on an SC-leu plate containing 0.25 mM CuSO_4 . **c**, Transcripts from the region of yeast chromosome V containing the *SNR14* gene. **d**, Nrd1-dependence of a 3'-end formation element in *SNR14* 3'-flanking region. Strains containing spontaneous suppressor mutations in the *SNR14* insert are shown in the bottom three rows. Cells were incubated at 30 °C for 3 days on SC-leu plates containing 0.7 mM CuSO_4 . **e**, Similarity of Nrd1-responsive regions from *SNR13* and *SNR14*. Numbering is relative to the transcription start site. Mutations that relieve Nrd1-dependent copper sensitivity are indicated by arrowheads. Asterisks denote a double-point mutation, and a trinucleotide deletion is underlined. Similarities to the U6R* artificial Nrd1-binding element^{7,8} and to a sequence in the Nrd1 mRNA 5' UTR (numbered relative to the Nrd1 start codon) are also shown.

sequence identity containing two consecutive CUU trinucleotides, one of which is targeted by the suppressor mutations. The same region of *SNR14* contains 7 out of 8 nt of the Nrd1-binding sequence present in the U6R* artificial Nrd1-dependent element^{7,8} (Fig. 3e). These *SNR13* and *SNR14* sequences may constitute RNA elements important for binding to Nrd1 and/or Nab3.

Notably, the *NRD1* 5' untranslated region (UTR) contains a 9-nt perfect match to the *SNR13* CUU repeat region (Fig. 3e), and the ORF array analysis indicates that Nrd1 mRNA increases in a *nrd1* mutant strain (N.K.C. and J.L.C., unpublished data). Northern blot analysis confirms that the abundance of Nrd1 mRNA is increased in a temperature-sensitive *nrd1* strain (Fig. 4a), and Nrd1 protein levels are coordinately increased in this mutant (N.K.C. and J.L.C., unpublished data). A similar response is also observed in temperature-sensitive *nab3* and *sen1* mutant strains, in which the wild-type Nrd1 message shows a rapid increase after temperature shift. Thus, regulation of Nrd1 mRNA levels and Nrd1-dependent 3'-end formation of snoRNAs and snRNAs use common factors, and probably a common mechanism.

Primer extension analysis mapped the principal *NRD1* transcription start sites to a region between 270 and 315 nt upstream of the AUG start codon (E.J.S. and D.A.B., unpublished data). We used a reporter construct comprising an *ADH1* promoter and the green fluorescent protein (GFP) coding region to test for a Nrd1-responsive element in the *NRD1* 5' UTR (Fig. 4b). Inserting the 5' UTR and a portion of the coding region from *NRD1* (bp -295 to +105 relative to AUG) downstream of the *ADH1* promoter and in frame with the GFP coding region results in a fourfold reduction in GFP mRNA in a *NRD1* strain (Fig. 4c, d). In contrast, the inserted sequence does not cause a reduction in GFP mRNA in a *nrd1* mutant strain (although the basal level of GFP message is reduced by about 50% in this strain). A smaller insertion lacking the 9-nt match to *SNR13* (bp -180 to +105) does not cause a reduction in GFP mRNA expression in either strain. These results indicate that a Nrd1-responsive element exists near the 5' end of the Nrd1 mRNA, where it would direct synthesis of truncated transcripts containing little or no Nrd1-coding sequence. This element elicits an efficient autoregulatory mechanism in which the rate of synthesis of full-length Nrd1 mRNA is inversely proportional to the amount of available Nrd1 protein.

Our results identify natural substrates of the essential Nrd1 pathway, including precursors of several snoRNAs, U4 snRNA and Nrd1 mRNA. Nrd1 function is required for normal maturation of at least some of the snoRNAs, including snR3, snR13 and snR190–snR128. Furthermore, Nrd1 apparently prevents transcription of snRNA and snoRNA genes from interfering with adjacent downstream genes. This function may be more vital to the cell than Nrd1-dependent 3'-end maturation of snoRNAs because most snoRNAs, including snR13, are dispensable for viability. *TRS31*, however, is an essential gene²³. Thus, interference with expression of *TRS31* and other genes adjacent to snoRNA transcription units may be lethal. For example, it is likely that mutations in *SEN1* inhibit pre-transfer RNA splicing⁹ not because Sen1 is directly involved in the process, but rather because consequent read through of the *SNR79* gene inhibits expression of the downstream *SEN2* gene (E.J.S. and D.A.B., unpublished data), which encodes a subunit of the tRNA splicing endonuclease²⁴.

The Nrd1-dependent pathway for 3'-end formation on non-polyadenylated transcripts shares several features with the pre-mRNA cleavage and polyadenylation pathway. Both pathways require recognition of sequence elements in the nascent transcript to direct RNA 3'-end maturation and to prevent pol II from reading through the normal termination region. In both cases, the protein complexes that recognize 3'-end formation signals interact with the pol II CTD. We previously noted that Pcf11, a component of the cleavage/polyadenylation factor CF IIA, contains sequence similarity to the CTD-binding domain found in Nrd1 (refs 7,8), and direct

interaction between Pcf11 and the CTD has been shown⁶. These similarities suggest the two pathways may share similar basic mechanisms for coupling termination to 3'-end maturation. Although components of the cleavage/polyadenylation pathway are well known, precisely how 3'-end formation of mRNA is coupled to transcription termination is not understood. The

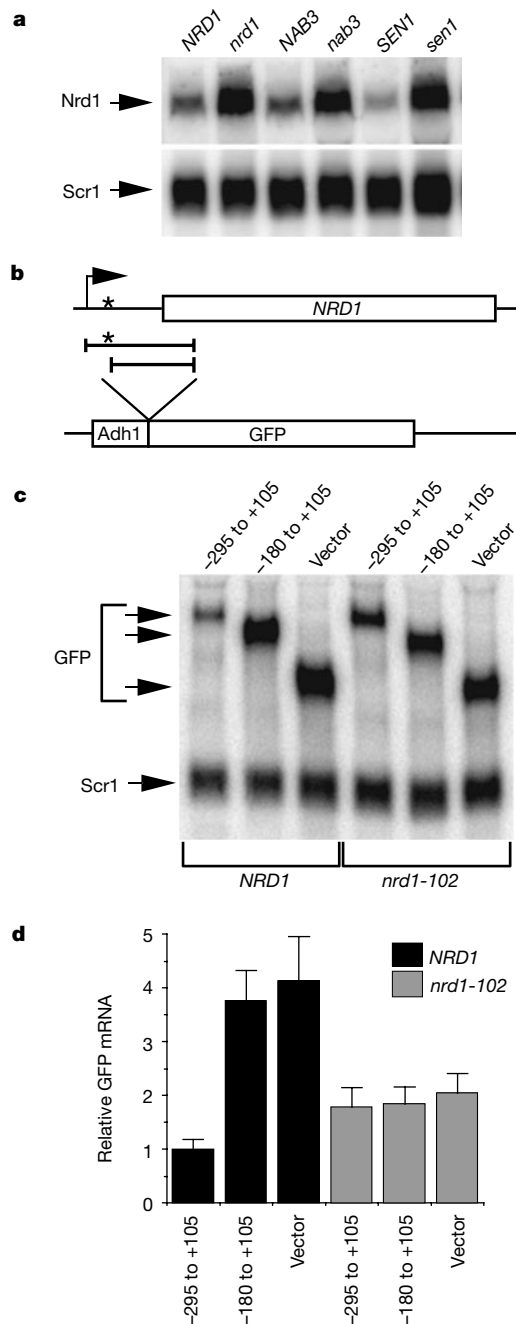


Figure 4 Autoregulation of Nrd1 mRNA accumulation by the Nrd1-dependent 3'-end formation pathway. **a**, Northern blot analysis of Nrd1 mRNA in the indicated temperature-sensitive mutant strains (and congenic wild-type strains) was performed with RNA extracted after a 2-h shift to 37 °C. Hybridization to Scr1 RNA serves as a loading control. **b**, Nrd1-GFP reporter constructs (see Methods for description). Asterisk indicates the approximate location of a 9-bp match to the *SNR13* Nrd1-responsive element. **c**, Northern blot of total RNA extracted from wild-type or temperature-sensitive *nrd1-102* mutant strains carrying the indicated GFP constructs. Samples were prepared and analysed as in **a**. **d**, Phosphorimager quantification of three independent Northern blot experiments. Values represent ratios of GFP to Scr1 signal, normalized to 1.0 for the -295 to +105 construct in the *NRD1* strain. Error bars represent s.d.

primary Nrd1-dependent event could be an endonucleolytic RNA cleavage that is linked to termination by the same (unknown) mechanism that couples termination to cleavage of pre-mRNA at the poly(A) site. Support for this possibility is provided by a recent study implicating cleavage/polyadenylation factors Rna14 and Rna15 in 3'-end maturation of snR13 (ref. 18). Further work should reveal the extent of overlap between the two 3'-end formation pathways. □

Methods

Yeast strains

The temperature-sensitive *nrd1* strains used in this study were *nrd1-5* (ref. 7) and YJC1166, which was generated by integrating *nrd1-102* (ref. 14) into the genomic *NRD1* locus of YJC582 (ref. 14) followed by counterselection of the plasmid-borne copy of *NRD1* on media containing 5-fluoroorotic acid (5-FOA). The *sen1* (*nrd2-1*), *nab3-11* and CTD truncation strains have been described^{7,14,25}. The *ctk1Δ* strain was generated by targeted disruption of the *CTK1* locus of strain YJC582 (ref. 14). The resulting strain was backcrossed to BY4741 (ref. 26) and sporulated, and *NRD1 ctk1Δ* spores were retained. Copper plate assays of *ACT-CUP* reporter gene activity were carried out in the *cup1Δ* strain 46a (ref. 19) with either *NRD1* or *nrd1-5* at the chromosomal locus. Sequences from *SNR13* and *SNR14* 3'-flanking regions were amplified by polymerase chain reaction (PCR) with primers containing *XhoI* sites, and inserted into the *XhoI* site of the *ACT-CUP* intron in pGAC24 (ref. 19). To generate the GFP reporter, the *Gall* promoter of pGal415 (ref. 27) was replaced with a PCR product encoding the *Adh1* promoter flanked by *SacI* and *XbaI* sites. The GFP coding region was similarly introduced through *SpeI* and *EcoRI* sites. Finally, a PCR product covering either the Nrd1 -295 to +105 or -180 to +105 region was inserted into the *SpeI* site with primer-encoded *XbaI* and *SpeI* sites.

RNA analysis

A more complete description of the expression profiling results will be published elsewhere (N.K.C., C. Colantuoni, J. Pevsner and J.L.C., manuscript in preparation). Briefly, log-phase cells from wild-type or temperature-sensitive strains were shifted to the non-permissive temperature for 45 min and total RNA was extracted. Radiolabelled cDNAs made by oligo-dT priming of total RNA were used to probe Yeast Index GeneFilters (Research Genetics, Huntsville, Alabama, USA). Data were analysed using the Pathways software (Research Genetics). Expression levels of ORFs *YJR129c*, *TRS31*, *YDR042c*, *YHR156c* and *YJL149w* increased markedly in the mutant strains and are located downstream of snoRNA genes *SNR3*, *SNR13*, *SNR47*, *SNR71* and the *SNR190-SNR128* dicistronic unit, respectively.

Northern blot analysis was performed with 20 µg of total RNA per lane (unless otherwise indicated). Probes were made by random priming of Klenow DNA polymerase on PCR products corresponding to the coding region of the gene of interest in the presence of [α -³²P]dCTP. For the blot shown in Fig. 1b, samples were run in duplicate, transferred to nylon membrane, and the membrane was split in half. Each half was probed with either a *SNR13* or *TRS31* coding region probe as indicated. For Fig. 1d, temperature-sensitive *nrd1* and *nab3* strains (and congenic wild-type strains) were shifted to 37 °C for 2 h, whereas CTD-26, CTD-10, *CTK1* and *ctk1Δ* RNAs were collected from log-phase cells grown at 25 °C.

We carried out reverse transcription as described⁷ with 10 µg RNA and ³²P-labelled oligonucleotides complementary to sequences located downstream of the mature 3' ends of the indicated snoRNAs. RNA was isolated after a 3-h (wild type, *nrd1-5*) or a 45-min (*sen1*) shift from 25 °C to 37 °C. An oligonucleotide complementary to the mature U6 RNA (a product of pol III) was used as an internal control in each experiment, and a representative example of the U6 reverse transcription products is shown. The oligonucleotide sequences and the sizes of expected products generated upon extension to the 5' end of the corresponding snoRNA are as follows: snR3: 5'-GATGCAAACTGTG-TCTCCTA-3', 458 nt; snR13: 5'-GACTCTGAGTATATTCTAGAGG-3', 549 nt; snR47: 5'-ACTCCAACACTATTTCGAGCA-3', 396 nt; snR71: 5'-GTTCCAGATGCTATAGAAGAGG-3', 689 nt; snR128: 5'-CCATTCCTCCAAAGTAGTGAT-3', 685 nt; snR33: 5'-CAGATAAACAAGCTCAGTAGTAA-3', 233 nt; snR39b: 5'-AACGTAATACGTAGAA-GAATGGA-3', 137 nt; snR45: 5'-ACAGATGAGATGACTACTCCCA-3', 223 nt; snR50: 5'-AAGAGTACTACTACGGGCTAGAA-3', 142 nt.

In vivo labelling

Cultures (10 ml) of the *nrd1-102* mutant and congenic wild-type strains were grown to mid-log phase in low-phosphate YEPD medium at 25 °C, and then shifted to 37 °C for 2 h before addition of 1 mCi of [³²P]orthophosphate. After labelling for 15 min at 37 °C, RNA was prepared and aliquots were hybridized to biotinylated oligonucleotides as described¹¹: SNR13-bio, 5'-bio-CCCCCCCCCAAGTAAAAAAGTAGCTTGAGT-3'; SNR3-bio, 5'-bio-CCCCCCCCCGGAAATTGATCTCTCCA-3'; SNR128-bio, 5'-bio-CCCCCCCCCTCTACCGTGGAAACTGCG-3'; SNR190-bio, 5'-bio-CCCCCCCCCTTGCTGTCATGGTCA-3'; biotin-6D, 5'-bio-AAACGAAA-TAAATCTCTTTG-3'. Hybrids were captured on streptavidin-paramagnetic particles (Promega) and washed, and then RNA was eluted into water essentially as described in the protocol supplied with the polyAtract system (Promega).

Transcription run-on analysis

We performed transcription run-on analysis essentially as described². Strains were shifted

to 37 °C for 2 h before run-on analysis. Single-stranded phagemid probes were isolated using standard laboratory methods. Phagemid constructs were generated by inserting PCR products into pBLUESCRIPT KS+ (Stratagene) using *SpeI* and *SacI* sites. Oligos used to generate PCR products were as follows: probe A, JCI363 (5'-AAGAGCTC-AACTTTGTCCTAAAGTACTA-3') and JCI323 (5'-GGACTAGTAGCAGCAACACGA-GATCCG-3'); probe B, JCI364 (5'-AAGAGCTCTTGTGCGGATAGGAGACAC-3') and JCI341 (5'-GGACTAGTAGGACAAGACGATGGCAAG-3'); probe C, JCI365 (5'-AA-GAGCTCGCAGCTTCCCATGCTCTGAA-3') and JCI366 (5'-GGACTAGTCAAATC-TAACGATGTCCTGC-3'). Relevant restriction sites are underlined. The actin probe has been described².

Received 27 April; accepted 26 July 2001.

- McCracken, S. *et al.* The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357–361 (1997).
- Birse, C. E., Minvielle-Sebastia, L., Lee, B. A., Keller, W. & Proudfoot, N. J. Coupling termination of transcription to messenger RNA maturation in yeast. *Science* **280**, 298–301 (1998).
- Hirose, Y. & Manley, J. L. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* **395**, 93–96 (1998).
- Osheim, Y. N., Proudfoot, N. J. & Beyer, A. L. EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage. *Mol. Cell* **3**, 379–387 (1999).
- Rodriguez, C. R. *et al.* Kin28, the TFIIF-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. *Mol. Cell. Biol.* **20**, 104–112 (2000).
- Barilla, D., Lee, B. A. & Proudfoot, N. J. Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **98**, 445–450 (2001).
- Steinmetz, E. J. & Brow, D. A. Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Mol. Cell. Biol.* **16**, 6993–7003 (1996).
- Steinmetz, E. J. & Brow, D. A. Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association. *Proc. Natl Acad. Sci. USA* **95**, 6699–6704 (1998).
- DeMarini, D. J., Winey, M., Ursic, D., Webb, F. & Culbertson, M. R. SEN1, a positive effector of tRNA-splicing endonuclease in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **12**, 2154–2164 (1992).
- Ursic, D., Himmel, K. L., Gurley, K. A., Webb, F. & Culbertson, M. R. The yeast SEN1 gene is required for the processing of diverse RNA classes. *Nucleic Acids Res.* **25**, 4778–4785 (1997).
- Rasmussen, T. P. & Culbertson, M. R. The putative nucleic acid helicase Sen1p is required for formation and stability of termini and for maximal rates of synthesis and levels of accumulation of small nuclear RNAs in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **18**, 6885–6896 (1998); erratum *ibid.* **19**, 5235 (1999).
- Gregor, I. H., Aranda, A. & Proudfoot, N. Balancing transcriptional interference and initiation on the GAL7 promoter of *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **97**, 8415–8420 (2000).
- Yuryev, A. *et al.* The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc. Natl Acad. Sci. USA* **93**, 6975–6980 (1996).
- Conrad, N. K. *et al.* A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. *Genetics* **154**, 557–571 (2000).
- Lee, J. M. & Greenleaf, A. L. CTD kinase large subunit is encoded by CTK1, a gene required for normal growth of *Saccharomyces cerevisiae*. *Gene Expr.* **1**, 149–167 (1991).
- Allmang, C. *et al.* Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J.* **18**, 5399–5410 (1999).
- van Hoof, A., Lennertz, P. & Parker, R. Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nuclear RNAs. *Mol. Cell. Biol.* **20**, 441–452 (2000).
- Fatica, A., Morlando, M. & Bozzoni, I. Yeast snoRNA accumulation relies on a cleavage-dependent/polyadenylation-independent 3'-processing apparatus. *EMBO J.* **19**, 6218–6229 (2000).
- Lesser, C. F. & Guthrie, C. Mutational analysis of pre-mRNA splicing in *Saccharomyces cerevisiae* using a sensitive new reporter gene, CUP1. *Genetics* **133**, 851–863 (1993).
- Chanfreau, G., Elela, S. A., Ares, M. Jr & Guthrie, C. Alternative 3'-end processing of U5 snRNA by RNase III. *Genes Dev.* **11**, 2741–2751 (1997).
- Abou Elela, S. & Ares, M. Jr Depletion of yeast RNase III blocks correct U2 3' end formation and results in polyadenylated but functional U2 snRNA. *EMBO J.* **17**, 3738–3746 (1998).
- Seipel, R. L., Zheng, B., Asuru, A. & Rymond, B. C. U1 snRNA is cleaved by RNase III and processed through an Sm site-dependent pathway. *Nucleic Acids Res.* **27**, 587–595 (1999).
- Sacher, M., Barrowman, J., Schieltz, D., Yates, J. R. III & Ferro-Novick, S. Identification and characterization of five new subunits of TRAPP. *Eur. J. Cell. Biol.* **79**, 71–80 (2000).
- Trotta, C. R. *et al.* The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. *Cell* **89**, 849–858 (1997).
- West, M. L. & Corden, J. L. Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* **140**, 1223–1233 (1995).
- Brachmann, C. B. *et al.* Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**, 115–132 (1998).
- Mumberg, D., Muller, R. & Funk, M. Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res.* **22**, 5767–5768 (1994).

Acknowledgements

We thank J. Dahlberg and J. Boeke for critically reviewing the manuscript, and members of the Brow, Corden and M. Culbertson labs for discussions. We also thank C. Colantuoni and J. Pevsner for assistance with the expression profiling experiments. This work was supported by grants from the NIH and the NSF.

Correspondence and requests for materials should be addressed to D.A.B. (e-mail: dabrow@facstaff.wisc.edu) or J.L.C. (e-mail: jorden@jhmi.edu).

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

US Head Office, New York

345 Park Avenue South, 10th Floor,
New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Director: Ben Crowe

US Sales Manager: Peyton Mason

US Advertising Coordinator:

Corissa Salzman

Classified Sales Representatives:

Phone: +1 800 989 7718

East USA/ NIH: Peter Bless

Mid West/ Corporate Advertising:

Frank Phelan

Corporate Advertising: Peyton Mason

East USA/ East Canada/

Latin America: Michael McGillion

Mid West US/ Physics:

Eric Wuerthenberg

San Francisco Office

Classified Sales Representative:

Michael Listek

West USA/ West Canada

225 Bush Street, Suite 1453

San Francisco, CA 94104

Tel +1 415 781 3803

Fax +1 415 781 3805

e-mail: m.listek@naturest.com

European Head Office, London

The Macmillan Building, 4 Crinan Street,

London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Group European Manager:

Leonie Weiss (4954)

European Manager:

Nevin Bayouni (4978)

Production Manager: Billie Franklin

To send films and materials use

London address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International Advertising

Coordinator: Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau

e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner

e-mails: p.phelan@nature.com

k.turner@nature.com

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Harakatsunomachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.cowan@naturejp.com

Japan Manager: Kate Cowan

Mergers and acquisitions

The marriage of cell biology and developmental biology is resulting in new scientific methods. In the process, it is also creating new institutes, perhaps most notably in Dresden, where the Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG) recently opened its doors.

Developmental biologists have generally concentrated on problems such as finding and understanding the genes that tell a young organism to sprout a wing rather than a leg. Cell biologists, for their part, have examined biomechanical processes such as cellular division and adhesion. Now, new technology and an increasingly integrationist approach to science are bringing the two fields together (see News Feature, page 244).

The MPI-CBG is one of the main institutes leading that approach (see *Naturejobs* this issue, pages 4–5). The institute's proponents are building up not just a new institute, but an entire biotech region on the emerging field's foundation. Apart from the MPI-CBG, the region will host a centre for biotech start-ups and an institute for computational biology. This construction has required the centres' leaders to marshal resources from the government, academia, industry and philanthropy. Ideally, as the science develops, fledgling companies and institutions built around it will expand, meaning more jobs.

Taking a leadership approach in a new field isn't easy. Expectations can run high — especially when the new frontiers are being bolstered by funds from both the public and private sectors. But the risks may be worth taking. Judging from the activity springing up around the Dresden institute, merging two disciplines can create net gains — both scientifically and economically.

Paul Smaglik

Naturejobs editor



Contents

REGIONS

Biopolis Dresden joins cell biology with its developmental cousin

p4

WWW.NATURE.COM/ NATUREJOBS

Career centre
Information on the scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



Rising star: the Max Planck Institute of Molecular Cell Biology and Genetics (above) has been galvanized by Kai Simons (above right).



Biopolis on the Elbe Dresden

It is reassuringly human to know that one person can make a difference. In Dresden, a city striving to become an academic powerhouse in the biological sciences, that person is Kai Simons, one of the directors of the Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG). “Kai has used his seniority in science to build biotechnology in the area,” says fellow director and long-term colleague Wieland Huttner. And as the founding president of the new European Life Scientist Organization and an expert in cell membranes, Simons has plenty of seniority to draw upon.

Simons’ aim, and that of a small founding coterie of scientists, is to create an open industrial network of academia and industry that will rival other world-class institutions. Together they are kick-starting three main engines of growth in an enterprise rapidly becoming known as Biopolis Dresden. These are the MPI-CBG, a new ‘BioTec’ centre (three-quarters industry, one-quarter academic) to be attached to the Technical University of Dresden (TUD), and the Centre for Computational Biology funded by the Klaus Tschira Foundation. Each is creating new jobs at all levels, from technician to professor.

A NEW INSTITUTE

The German government turned its attention to Dresden in the early 1990s, as part of a plan to spread centres of intellectual excellence into the former East Germany. But “they were having trouble recruiting someone of calibre to get the enterprise off the ground”, says Huttner.

This was not surprising, argues Simons, given how competitive molecular biology and genetics are, and

how exciting environments such as Harvard and Britain’s Cambridge are with their mixture of academia, start-ups, universities and related institutes. In 1996, Huttner told the Max Planck Society (MPS) that a traditional research-only institute would not work. Something more was needed and Simons, he suggested, was the person to build it.

That something was a whole new infrastructure that could form a biotechnology network, with open interactions between universities, institutes and industry, says Simons. The crucial question was: “Is it possible to establish a world-class research centre where it did not exist before and be taken seriously by other world-class centres?” Simons says.

In the meantime, the state government in Saxony had concluded that biotechnology would be a welcome addition to the area’s economy. The state was also aware that the MPS was looking for a location for a biologically based institute, but it too — as the provider of half of the funds — wanted more than a research-only institute.

Aware of all these swirling aspirations, Simons said yes to the overtures of the MPS in 1996. A little over a year later, the MPI-CBG was founded, and the institute opened its doors for business at the end of January this year. The price tag for construction and equipment was DM130 million (US\$59 million), and it now employs around 200 people and has space eventually for about 400. The MPI-CBG’s focus is the

Right: Klaus Tschira is laying the foundations for a bioinformatics centre on the banks of the Elbe.



interface between cell and developmental biology (see News Feature on pages 244–246). Of the 25 group leaders who will eventually staff the institute, 20 have so far been recruited — one of whom will work from the International Institute of Molecular and Cell Biology in Warsaw.

Huttner heads the international PhD programme at the MPI-CBG, which covers cell biology, developmental biology, bioengineering, genetics, biophysics and neurobiology. Some 30 graduates were selected in 2000–01, and this number will rise to 120 by 2003–04.

The MPS's rule for its international programmes is that no more than half of the fellowships can go to Germans, Huttner explains. In the last round, he was particularly pleased to see a high percentage of high-quality candidates from eastern and central Europe.

INDUSTRIAL AND ACADEMIC PARTNERSHIPS

So much for the institute, but what about the 'something more'? Through 1998 and 1999, while the institute was building up its research groups, Simons and colleagues, including Konrad Müller, a consultant and former head of personnel at the European Molecular Biology Laboratory, immersed themselves in the business of giving Dresden the infrastructure of a world-class centre of science.

They worked with the TUD and the state government to establish the intellectual framework and aims of a new centre that would bring biology, medicine and engineering together. The result is a 15-acre site, part of the TUD, in the centre of Dresden and close to the MPI-CBG. The building site will provide a home for industry and academia to work side by side. Such proximity and open interchange between academia and industry is essential for the intellectual health of biology in the area, argues Simons, because the two feed off one another's ideas.

Start-up and young companies will eventually rent three-quarters of the centre, and have access to start-up advice, public relations, marketing, workshops and seminars, as well as central scientific facilities that academics and industrialists will share.

Academics will occupy the remaining quarter. The TUD and the University of Leipzig have created five new professorships covering genomics, biophysics, proteomics, tissue engineering and cell machinery at the new institute. And it is Simons and his colleagues that provide the pulling power to bring in talented researchers to fill these posts.

The centre will be completed in 2003. In the meantime, both academics and start-ups are occupying spare space in the MPI-CBG (see 'The path from lab to industry', right).

PHILANTHROPIC SUPPORT

In parallel with the administration and planning needed for both the MPI-CBG and the BioTec centre, Simons and his colleagues were strongly aware of the need for expertise in bioinformatics. "We put feelers out for money to private institutions, the city and the European Union," says Simons.

They struck lucky with the Klaus Tschira Foundation, a charitable organization established to promote the natural sciences and protect Germany's architectural heritage. Klaus Tschira is the co-founder

The path from lab to industry

Dresden's biopolis has had a radical effect on the career of Chris Echeverri (right). Once an academic, he is now chief executive of Cenix BioScience, a biotech start-up he co-founded in 1999. Cenix is currently based at the Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG), and will move into the new BioTec centre at the Technical University of Dresden when it opens in 2003.

Cenix arose from research Echeverri did at the European Molecular Biology Laboratory in Heidelberg. With Pierre Gönczy, he scaled up a lab technique for identifying the genes involved in cell division. Such genes can be useful for finding potential new targets for anticancer drugs.

The team worked on one chromosome in the nematode worm *Caenorhabditis elegans* (see *Nature* 408, 331–336; 2000). But analysing the whole genome would have taken over the entire lab, says Echeverri. So starting a company



seemed like a good idea.

Echeverri found the decision to leave academia very hard. "I probably wouldn't have made the move if it had been to an existing company, but this was an opportunity to do something new," he says. "I was craving exposure to a broader range of biology."

In fact, by leaving academia, Echeverri was exposed to far more. "The biotechnology business is a completely different world from academia," he says. "It has its own psychology, jargon and dynamics — there is no way pure research scientists can figure the biotech business out without help."

H.G.

of software company SAP and is generally viewed as Germany's Bill Gates. His foundation is restoring Lingner castle on the opposite bank of the Elbe from the MPI-CBG. This will provide space for start-up companies and meeting rooms. The foundation is building a bioinformatics centre on the grounds, and another new professor, created by the TUD and the University of Leipzig, will head the centre.

It is an exciting time, says Tony Hyman, one of the four additional directors that Simons tempted to join him at the MPI-CBG. "All the balls are still in the air," he says, "they're still coming down. So far we haven't dropped any." Providing they don't, it seems that the vision of a world-class centre of biology in Dresden could succeed.

Helen Gavaghan is a freelance writer based in West Yorkshire.

Max Planck Institute of Molecular Cell Biology and Genetics

► <http://www.mpi-cbg.de>

Technical University of Dresden ► <http://www.tu-dresden.de>

Cenix BioScience ► <http://www.cenix-bioscience.com>

European Life Scientist Organization ► <http://www.elseo.org>

Bright future: Tony Hyman knows it is something of a balancing act, but believes Dresden's biopolis will be successful.

